THIS WEEK

The innovation game

Innovation within the European Union is wanting for reasons cultural, historical and technical. It can best be strengthened by breaking down barriers and building a united research area.

A week is a long time in politics, as one-time British prime minister Harold Wilson famously said. But in European Union (EU) politics, a decade can seem very short indeed.

Just look at the ten-year strategic plan for economic growth and improved welfare that EU heads of state signed up to in Lisbon in 2000, in which research had a central role. The three EU bodies — the Council, Parliament and Commission — each realized the urgent need to make Europe work as a single territory for scientists, rather than separate bordered countries — now numbering 27 — with their own languages and habits. They agreed to create the European Research Area, intended to free the movement of scientists between countries by breaking down barriers such as difficulties in transferring pensions or transporting national research grants. They endorsed the concept of a single patent that would be valid EU-wide. And they agreed on a target to spend 3% of gross domestic product on research and development by 2010.

But ten years didn't prove long enough to achieve these aims. Once home, national governments were unwilling to concede sufficient sovereignty. The European patent, for example, depends on an agreement to work in a limited number of languages to keep patenting costs reasonable — but several countries still insist that all documents be translated into their own languages. Others want to protect the revenues of their national patent shops. Little headway has been made towards the legislative changes in areas such as pensions that were required to build the European Research Area. And most nations have failed to significantly increase their public research spending, or to incentivize that of the private sector.

Fortunately, the European Commission has stuck to each of these fundamental goals in its latest proposal for a research-related strategy for the next decade, which was released on 6 October. Called the Innovation Union, the new strategy is a component of the Lisbon Agenda's successor, Europe 2020, which was launched in March (see *Nature* **464,** 142; 2010). The EU Competitiveness Council, which comprises national research and industry ministers, is now preparing a response to the Innovation Union document, which will be discussed by the heads of state at a summit meeting on 16 December.

The Commission is dead right to persist with the research objectives of the Lisbon Agenda, because until these are achieved, Europe will not be able to compete. It is also right to emphasize the role of the European Investment Bank in providing much-needed cross-border risk capital, which is barely available in Europe.

Less convincing, unfortunately, is its fresh proposal for what it calls 'innovation partnerships' — elaborate-sounding efforts to engineer alliances between everyone in the innovation chain, all the way from researchers and manufacturers to consumer representatives, to tackle big societal problems. These partnerships will focus on a set of established 'grand challenges', such as the ageing society, climate change and food security. The first of the new partnerships will address 'healthy ageing', the Commission suggests.

If this sort of approach sounds familiar, that's because a number of related ones are already under way. Within one called 'joint programming', for example, national research efforts are supposed to be coordinated independently of the Commission. Another idea, for 'joint technology initiatives', set up public–private research partnerships, co-funded by the Commission. And the European Institute of Innovation and Technology has morphed into another series of public–private partnerships called Knowledge and Innovation Communities.

None of these initiatives can yet be considered successful — they are in their infancy and still being fine-tuned. The innovation partnerships will perpetuate — and further complicate — the tradition, and even aim to tap into public services and their budgets, which are unfamiliar territory for EU research partnerships.

The Healthy Ageing innovation partnership has the remarkably ambitious target of yielding a two-year increase in the age to which the average EU citizen enjoys good health, by 2020. The target is laudable and simple. But is the general strategy correct? It may take many more years to create the European Research Area, but this is really what matters. In the meantime it would be best to get existing initiatives to work better before adding new ones. Once the legislative problems are solved, and risk-capital mechanisms in place, innovation should emerge on its own — without having to engineer it. ■

> *The European Commission is right to emphasize the role of the European Investment Bank.*

Not quite assured

An upbeat assessment of phosphate reserves leaves several questions unanswered.

Phosphorus in the form of phosphate has a crucial involvement in RNA, DNA and cellular metabolism, and all forms of life depend on it. Along with nitrogen and potassium, phosphorus is essential for healthy plant growth — and its supply through fertilizer is a mainstay of modern agriculture.

Reserves of the phosphate rock used to make such fertilizers are finite, and concerns have been raised that they are in danger of exhaustion. It has been argued, for example, that data from the US Geological Survey point to the available supplies peaking in as little as 25 years time (see *Nature* **461,** 716–718; 2009). Because there is no substitute for phosphate in agriculture, this might present an

urgent and substantial problem. But initial findings from the *World Phosphate Rock Reserves and Resources* study conducted this year by the IFDC, an international non-profit organization based in Muscle Shoals, Alabama, and formerly known as the International Fertilizer Development Center, suggest that phosphate rock deposits should last for between 300 and 400 years.

Accurate information about phosphate reserves is hard to come by, and the IFDC concedes that more work is needed to hone its estimates. The mining industry, governments and interested researchers should accept the organization's invitation to collaborate in this process.

The phosphate issue runs beyond gaining assurances that total global supply will meet demand. There remain important concerns that phosphate and other fertilizers are being squandered in some parts of the world, whereas farmers in other regions cannot obtain them at a reasonable cost.

After decades of wanton overuse, farmers in the United States, Europe and elsewhere are now using sophisticated assessments to tell them when, how much and in what proportion fertilizer should be applied. That has led to a flattening out in global demand for phosphate fertilizer, despite continued growth in food production.

But elsewhere in the world, especially in Asia, farmers are still applying fertilizer in excess (see *Nature* doi:10.1038/news.2010.498; 2010). At the same time, farmers in the poorest countries such as some in Africa, find fertilizer prices inflated to unaffordable levels by high transportation costs and local market conditions.

In addition, current fertilizer-production methods fail to maximize the efficient conversion of phosphate rock into fertilizer. The supply of the rock is heavily concentrated in two nations, China and Morocco, on whose good faith the rest of the world relies for its phosphate supplies. That faith has been shaken by extreme price fluctuations in recent years.

Yet the heavy dependence of food production on fertilizers, inequalities of supply and the need for sustainable use of fertilizers — including recycling — are largely missing from discussions on approaches to sustainable development. They were only mentioned in passing, for example, at the United Nations' world summit on food security in Rome last November.

> *"The need for sustainable use of fertilizers is largely missing from discussions on approaches to sustainable development."*

Hydrologists, soil researchers and food scientists have begun to raise awareness of some of the issues surrounding phosphates. A discussion will be devoted to the topic at the Crop World 2010 meeting in London next week, in which researchers will be joined by industry and government representatives, including John Beddington, the UK government's chief scientific adviser, who has worked hard to raise political awareness of food-security issues.

These efforts would be strengthened if an international body, such as the UN Food and Agriculture Organization, started to seriously champion the issue of sustainable fertilizer use. The organization already tracks fertilizer demand and supply, and has produced reports on phosphate fertilizer use. It doesn't have a specific programme for sustainable fertilizers, but its departments of agriculture and natural resources do some work in this area, giving it a base on which to build. It now needs to push this issue out from the sidelines and into the policy-making process that will shape the future of agriculture and sustainable development. ∎

Space hitch–hiker

Commercial spacecraft with room to carry experiments could give science a lift.

A study on the environmental impacts of space tourism suggests that a surge in private access to space could speed global warming. Led by Martin Ross, an atmospheric scientist at the Aerospace Corporation in El Segundo, California, it shows that sooty emissions from 1,000 rocket launches per year would add as much to climate change as current emissions from the global aviation industry. It has been accepted for publication by *Geophysical Research Letters*.

Perhaps the most striking aspect of the study is not the projected impact on polar temperature and sea ice, but the size of the industry it models. Three launches a day? Don't bet against it. Barely a decade after US multimillionaire Dennis Tito paid around US$20 million for a trip to the International Space Station (ISS), space tourism, at least the suborbital type, seems poised for serious lift-off.

The private spaceflight industry is making steady progress. Spaceport America, a launch site in Las Cruces, New Mexico, opened its first runway last week. Earlier this month, US President Barack Obama signed into law the NASA Authorization Act, which, subject to approval by Congress, will see the agency hand over $15 million a year to help commercial suborbital efforts.

NASA is keen because it sees what many space scientists have been slow to realize: such suborbital flights could carry research payloads. Virgin Galactic, a pioneer of space tourism, has already indicated that it would be happy to host scientific experiments on its SpaceShipTwo vehicle. A number of fields including atmospheric, space and microgravity research could benefit. A closer relationship with scientists could help the industry in return,

➦ **NATURE.COM**
To comment online, click on Editorials at:
go.nature.com/xhunqv

through work to quantify and reduce its environmental impact, for instance.

A strong advocate of closer ties between rocketeers and researchers is Alan Stern, a planetary scientist at the Southwest Research Institute in San Antonio, Texas, and a former NASA associate administrator, who chairs the Suborbital Applications Researchers Group of the Commercial Spaceflight Federation in Washington DC. Stern says that private suborbital vehicles will be a game-changer for science, because of low costs and the high number of flights. Earlier this year, his group organized the first conference to promote the benefits of private space flights to scientists. A second event is scheduled for February 2011 at the University of Central Florida in Orlando.

Space scientists who wish to fly experiments currently face high costs and long waits for room on the ISS or sounding rockets, or frustratingly brief periods of microgravity in drop-tubes or parabolic aircraft (known with little affection by those who have been aboard as 'vomit comets'). Suborbital flights could offer several minutes of weightlessness for a fraction of the cost of a conventional launch. And the experiments could be supervised by scientists able to fly alongside their kit. An early winner could be the search for vulcanoids — asteroids that orbit the Sun closer than Mercury. None has yet been discovered, perhaps because observing them from the ground or high-altitude flights is so awkward.

Although NASA has been quick to identify and nurture the potential of space-tourism operators, others have been more sluggish to recognize their potential. The European Space Agency, for example, has an official position on private suborbital flights only of "cautious interest and informed support". Countries outside the United States have not yet taken the necessary legal steps to open their skies to private operators. Perhaps this reflects scepticism about whether the endeavour will reach the necessary economy of scale, which depends on the number of tourists who sign up. That is a reasonable position at this stage, but space scientists and administrators should drop any snobbish objections they have to the private sector. Those who do not embrace the possibilities could find themselves, quite literally, left behind. ∎

Spending review leaves research in the lurch

A revised research spending plan won't meet the challenges Britain faces from its international competitors or from climate change, argues **David King**.

Last week, the UK government announced its plans for cutting an astonishing £81 billion (US$128 billion) from the country's budget over the next four years. Although other departments saw an average of 19% shorn off their annual funding, science got off relatively lightly. The United Kingdom's research budget was frozen but not cut, meaning an effective reduction of some 10–12%.

That sounds like good news, but there are two main problems. First, we do not yet know the indirect effect that the cuts to university teaching budgets will have on research, nor how much they can be offset by increased student fees. Second, and perhaps more important for the research community, because the funding for large international collaborations such as CERN, Europe's particle-physics laboratory near Geneva, Switzerland, has to be ring-fenced, most of the cuts will fall on shorter-term, more timely pieces of research.

This means that certain research councils face a far larger percentage cut. The Engineering and Physical Sciences Research Council, for example, has few long-term commitments, so only a small part of its budget is ring-fenced. The rest will be fair game to meet not just its own share of the overall target, but also that of councils with larger ring-fenced allocations. There could even be funding rounds in which it is unable to allocate any grants at all. This in turn means that timely ideas could fall by the wayside, or be taken up by international competitors.

This matters, to Britain at least, because I believe that research funding lies at the heart of the country's economic recovery and future prosperity. In 2000, the UK government that I advised realized that, in the following decades, science and technology — and the innovation and wealth creation that follows — would be more in demand than ever before. Humanity faces unprecedented challenges: the deterioration of ecosystems; resource mismanagement and shortages; and decarbonizing the economy, which is the biggest single innovation challenge since the Industrial Revolution.

For these reasons and more, the ten-year strategy setting out the previous government's science and investment framework for 2004–14 pledged to continue to increase the science budget each year by twice the rate of growth in gross domestic product (GDP) (but not to reduce the budget if GDP contracted, as it has done recently). `

This made waves around the world — notably in the emerging markets that are providing Europe and the United States with an increasing (and I would say, welcome) economic challenge. In 2003, the Chinese premier Wen Jiabao asked to meet me during a state visit. Why? Because the prime minister's 2002 speech 'Science Matters' had been translated into Chinese and he wanted to know more. When I went to China the following year, the Chinese government declared that it had decided to match the UK pledge of increasing science funding by twice the level of GDP growth. But China committed to doing this over 20 years, not 10, and as its GDP growth was 10%, it has been boosting its science budget accordingly — with a 30% increase from 2008 to 2009. Even this year it has continued the increase, with an 8% rise in the science budget. This is underpinning the nation's continuing remarkable economic growth and the increased competitiveness of its manufacturing industry.

The United States, too, has seen the need for change. The administration of President Barack Obama has revitalized US research through public funding over the past year, substantially increasing research funding across the board, as well as giving a large boost to alternative-energy research (see *Nature* doi:10.1038/news.2009.457; 2009).

Europe is also focusing on research funding. In May, leaders in European research, industry and policy met under the aegis of the European Research Area Board, of which I am a member, to consider the European Union's research, development and innovation policy. Its report calls for radical action, including the establishment of a single market for research and development. And in the past few months, both France and Germany have published national strategies showing their commitment to investing in research.

So, although the cut in the UK science budget is lighter than I had feared, I still believe that it threatens the country's ability to use the power of science research to retain its international competitiveness. Just as importantly, it threatens the country's ability to decarbonize the economy. Most of the funding for Britain's energy research comes through the research councils, and it is deeply worrying that this will be cut just when a radical increase in activity is needed. Admittedly, there was some good news in this regard, as the government reinforced its funding for energy and the environment in the Department of Energy and Climate Change and the Department for Environment, Food and Rural Affairs. This will be crucial if Britain is to stick to its commitment of reducing carbon dioxide emissions by 34% by 2020.

However, the agenda set out by the UK government in 2004 in its ten-year strategy for research was always intended to be a long-term investment. The danger of the freeze proposed by the present government is that it could stall the whole process just as it is taking off. In the meantime, watch out for a bloodbath as scarce resources are divided between the research councils this winter. ∎

> THE CUTS THREATEN THE COUNTRY'S ABILITY TO USE **THE POWER OF SCIENCE** TO RETAIN ITS INTERNATIONAL COMPETITIVENESS.

David King *was chief scientific adviser to the UK government from 2000 to 2007, and is now director of the Smith School of Enterprise and the Environment at the University of Oxford.*
e-mail: director@smithschool.ox.ac.uk

Bacterial cyborg transmits electrons

One idea for biosensors and bioenergy is to combine living cells with inorganic materials. Researchers have taken a step towards this goal by engineering the bacterium *Escherichia coli* to transmit electrons to inorganic materials.

Cell membranes act as insulators and thus hinder the movement of electrons between cells and inanimate materials. Caroline Ajo-Franklin at the Lawrence Berkeley National Laboratory in Berkeley, California, and her colleagues overcame this by introducing genes for electron-shuttling proteins into *E. coli*. The genes occur naturally in another bacterium, *Shewanella oneidensis*, which can transfer charge to non-living materials in oxygen-free environments.

The engineered *E. coli* cells were able to reduce iron in culture six to eight times faster than normal strains. The authors say that these genes could be transferred to other microbes to create, for example, low-cost photobatteries — by inserting them into bacteria that generate electrons in response to light.
Proc. Natl Acad. Sci. USA doi:10.1073/pnas.1009645107 (2010)

F. LANTING/FLPA

H. MONTGOMERY/AP PHOTO

Better fishing for the future

Despite European Union rules controlling fishing catch sizes, fish stocks are collapsing. Change is needed to maintain populations at levels that can produce maximum sustainable yields, according to Rainer Froese at the Leibniz Institute of Marine Sciences in Kiel, Germany, and his colleagues. They have devised new rules that take a more cautious approach: limiting catches to levels that would leave species biomass at 1.3 times the total needed to produce maximum sustainable yields.

The current system, which regulates catch sizes according to the size of the smallest fish stock that could still deliver sustainable catches, encourages overfishing, the authors add. They say that their proposed rules would have prevented the collapse of the North Sea herring (*Clupea harengus*) in the 1970s.
Fish Fisheries doi:10.1111/j.1467-2979.2010.00387.x (2010)

Leopards change their spots

Tree-living cats that hunt by night in dense environments tend to have more complex coat patterns than plains-dwelling felines that are active during the day. The patterns seem to evolve relatively rapidly in response to environmental change and help the animals to remain camouflaged.

William Allen and his colleagues at the University of Bristol, UK, analysed images of coat patterns in 35 cat species, including leopards, jaguars (**pictured**) and tigers. They used a mathematical model to link pattern development and function to habitat and behavioural traits.

They also mapped pattern variation on a felid family tree. This revealed that patterns have changed frequently during felid evolution, suggesting that coat pattern is under simple genetic control.
Proc. R. Soc. B doi:10.1098/rspb.2010.1734 (2010)

Quiescent cells not so quiet

Many of the body's cell types enter a state in which they do not divide and, or so scientists thought, reduce their metabolic rates. But Hilary Coller and her colleagues at Princeton University in New Jersey show that quiescent human fibroblast cells — common in connective tissues — have similar metabolic activity to their proliferating counterparts.

The team measured and analysed the levels of 62 metabolites extracted from the cells, as well as levels of secreted proteins. They found that quiescent cells were busy breaking down and resynthesizing proteins and lipids, as well as secreting proteins that help to maintain tissues. Moreover, inhibiting a metabolic pathway in these cells led to increased programmed cell death, leading the authors to suggest that certain dormant cells, such as cancer stem cells, can be selectively killed.
PLoS Biol. 8, e1000514 (2010)

CHEMISTRY

The hunt for explosives

A dye-based sensor can detect tiny amounts of an explosive that has been used in several terrorist incidents.

Current methods for detecting triacetone triperoxide (TATP) have several drawbacks, such as being cumbersome or expensive. Kenneth Suslick and Hengwei Lin at the University of Illinois at Urbana-Champaign have developed a way to sense TATP levels as low as 2 parts per billion. They show that, in a gas flow treated with a solid acid catalyst, TATP decomposes into products, such as hydrogen peroxide, that can be detected with a colorimetric sensor.

The researchers have created a prototype hand-held detector that could be used to screen luggage. Importantly, the detector is not activated by other common compounds such as soaps, liquors or volatile organics.

J. Am. Chem. Soc. doi:10.1021/ja107419t (2010)

PHYSICS

Insulator insight into constant

Enigmatic materials that conduct electricity at only their surfaces, known as topological insulators, could be used to measure the fine structure constant, α — one of three factors that determine the speed of light.

Joseph Maciejko at Stanford University in California and his team propose measuring α by observing the quantized magnetoelectric effect (QME) — a predicted phenomenon in which an electric field induces magnetization in discrete quantum steps. The proposed experiment would use a layer of a topological insulator on top of a layer of ordinary insulator, all sitting in an external magnetic field. The authors say that measuring the polarization of light reflected off the surface of the topological insulator, and comparing this with the measurement of the polarization of light transmitted through the layers, will reveal a measurement of the QME — and hence α — in a way that is independent of the materials' properties.

Phys. Rev. Lett. **105**, 166803 (2010)

CELL BIOLOGY

Spindle-free division in yeast

During cell division, or mitosis, protein microtubules called spindles pull the replicated chromosomes apart before the cell splits in two. Stefania Castagnetti at Cancer Research UK in London and her group show that some yeast missing these spindles undergo a novel form of nuclear division — which may be a primitive form of mitosis.

Schizosaccharomyces pombe strains treated with a chemical that breaks down microtubules could still separate their chromosomes. By probing individual parts of the mitotic apparatus, the researchers surmise that, in the absence of spindles, the chromosomes remain associated with the cell's two spindle pole bodies, which normally act as anchors for the spindles. The authors suggest that these organelles move away from each other within the nuclear membrane, carrying the chromosomes along with them, before the nucleus divides.

PLoS Biol. **8**, e1000512 (2010)

ECOLOGY

What mammoths left behind

Mass extinction of most of the world's large mammals some 10,000 years ago liberated roughly 1.4 petagrams of plant life previously consumed as food. The surplus endured until human populations grew to fill the void.

Christopher Doughty, now at the University of Oxford, UK, and Christopher Field at the Carnegie Institution in Stanford, California, estimated consumption by the extinct Pleistocene megafauna and by humans, and compared the results with net primary plant production around the globe. Averaging the figures out worldwide, they found that liberated plant resources — about 2.5% of net terrestrial productivity — had been used up by humans by about 1700.

The duo also showed that by 2000, humans were consuming roughly six times more than the megafauna had done. Meanwhile, human agriculture had reduced global primary productivity by about 10% as a result of factors such as land degradation.

Environ. Res. Lett. **5**, 044001 (2010)

BIOCHEMISTRY

Zooming in on proteins

High-resolution optical imaging of single molecules has been achieved in living cells through the design of a small fluorescent organic molecule that outperforms commonly used fluorescent proteins.

Organic fluorophores generally emit much more light than fluorescent proteins.

MOLECULAR BIOLOGY

Long RNAs turn up gene expression

Long RNA molecules that do not code for proteins boost the expression of certain human genes, including those linked to development. Typically, regulatory RNAs, such as microRNAs, quiet gene expression.

Ramin Shiekhattar at the Wistar Institute in Philadelphia, Pennsylvania, and his colleagues found 3,019 RNA molecules, averaging 800 nucleotides in length, after scouring a portion of the human genome. When the team stimulated the development of a type of human skin cell, expression levels of many of the long non-coding RNA molecules rose in step with those of nearby protein-coding genes. Reducing the levels of a set of the RNA molecules in various cell lines also curbed the expression of neighbouring genes, including one coding for a protein that regulates blood-cell development.

Cell **143**, 46–58 (2010)

William Moerner of Stanford University in California and his colleagues have devised a system in which a commercial enzyme is fused to the protein of interest. The 'fluorogen' then binds to the enzyme and is activated by light, enabling high-resolution imaging by the controlled activation of single molecules.
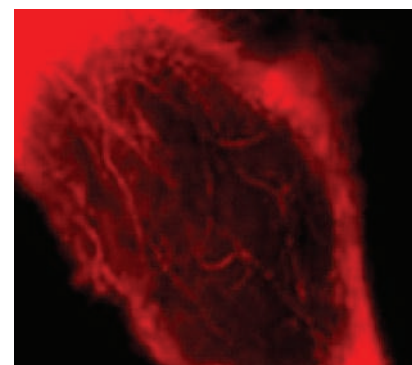
The authors were able to image protein microtubules in mammalian cells (**pictured**), as well as other protein structures in living bacteria, with a resolution beyond the limit of optical diffraction.

J. Am. Chem. Soc. doi:10.1021/ja1044192 (2010)

AM. CHEM. SOC.

↻ **NATURE.COM**
For the latest research published by *Nature* visit:
www.nature.com/latestresearch

Scientific integrity

The White House's Office of Science and Technology Policy (OSTP) is in court over its failure to put forward recommendations to ensure scientific integrity in government. Scientists are still waiting, 18 months after President Barack Obama gave the OSTP 90 days to deliver agency guidelines for putting science at the centre of policy-making. Public Employees for Environmental Responsibility, an advocacy group based in Washington DC, wants to know why. It started legal action against the OSTP on 19 October when the agency didn't respond to its freedom of information request for draft recommendations. See go.nature.com/aec5zz for more.

Misconduct report

A panel commissioned by the Canadian government has recommended that the nation revise its system for curbing research misconduct. A 21 October report by the Council of Canadian Academies — a non-profit organization based in Ottawa — says that a council of research integrity should be created to help educate researchers about good practice and to provide confidential advice. Privacy laws hampering the identification of individuals or institutions found guilty of research misconduct should also be relaxed, the report says. See go.nature.com/ISyJDi for more.

Science-prize row

The United Nations Educational, Scientific and Cultural Organization (UNESCO) last week found a way to avoid awarding a controversial science prize sponsored by an African dictator, whose regime is

Conservation's rare successes

Fifty-two species of vertebrates move a category closer to extinction every year, according to an analysis of more than 25,000 mammals, birds and amphibians published on 26 October (M. Hoffman *et al. Science* doi:10.1126/science.1194442; 2010), as the parties to the Convention on Biological Diversity meet in Nagoya, Japan. But falling biodiversity has been slowed by conservation efforts, such as those that repopulated parts of North America with the still-endangered black-footed ferret (*Mustela nigripes*, pictured). Using an index of extinction risk based on category movements in the International Union for Conservation of Nature (IUCN) Red List, Michael Hoffman at the IUCN in Cambridge, UK, and his colleagues found that biodiversity declines would have been at least one-fifth worse without any efforts to halt habitat loss, curb hunting and tackle invasive species. The last was the most effective strategy, they said.

widely viewed as corrupt and oppressive. The Paris-based organization is not explicitly rejecting the life-sciences prize, funded by a US$3-million donation from President Teodoro Obiang Nguema Mbasogo of Equatorial Guinea. Instead, the UNESCO executive board agreed to suspend awarding the money "until a consensus is reached" — a diplomatic way of putting the prize on hold, as it seems unlikely that delegate nations will ever agree. See go.nature.com/Nbi9nQ for more.

UK funding

British scientists were jubilant at escaping the worst of government cuts announced on 20 October.

The science budget was frozen at £4.6 billion (US$7.2 billion) annually for four years — although other government departments saw spending drop by an average of 19%. See page 1017 for more.

Stem-cell funding

The California Institute for Regenerative Medicine (CIRM) announced awards on 21 October worth US$72 million to fund 19 stem-cell researchers in the state, as well as to recruit another. Last year, the agency funded 14 researchers with $230 million; the grants are aimed at moving experimental treatments into the clinic. Funded by a $3-billion bond in 2004, CIRM has $1.6 billion remaining in its coffers. Meanwhile, a US Court of Appeals will hear arguments

in a lawsuit next month challenging the National Institutes of Health's ability to fund human embryonic stem-cell research. See nature.com/stemcellfunding and page 1031 for more.

Cholera in Haiti

More than 250 people have died from the cholera outbreak in earthquake-ravaged Haiti, the United Nations said on 25 October. Some 3,000 people have contracted the disease, which spreads through contaminated water and food. Although cholera claims thousands of lives in African countries every year, it is Haiti's first outbreak in a century. As *Nature* went to press, aid workers hoped

that the outbreak could be prevented from spreading in the capital Port-au-Prince. For more analysis of the Haiti earthquake, see page 1018.

RESEARCH

Volcano drilling

A project to drill a borehole into an active volcano near Naples, Italy, has been halted awaiting further safety data. Researchers at Italy's National Institute for Geophysics and Volcanology (INGV) in Naples had planned to drill 4,000 metres into the Campi Flegrei volcano to learn what signs might precede an eruption. But some Italian scientists voiced concerns about health and environmental risks (see go.nature.com/eH4FEV). On 18 October, the mayor of Naples, Rosa Russo Iervolino, said she had asked the Italian civil-protection department for a safety report, which is likely to take a few weeks. INGV scientists say the project is safe.

BUSINESS

Rare earth alarm

A simmering trade dispute over rare earth elements intensified last week, as Japan urged China to resume exporting the minerals; it says shipments have been blocked since September, although Beijing denies an official export ban. Meanwhile, share prices of rare-earth mining companies continue to rocket, and US congressman Ed Markey (Democrat, Massachusetts) has asked the US government to look into reports of additional Chinese export curbs. Miners in China (pictured) produce more than 90% of the world's rare earth elements, which are used as catalysts and in high-tech magnets, car batteries, wind turbines and mobile phones.

Obesity drugs

The US Food and Drug Administration (FDA) maintained its cautious approach to weight-loss drugs on 23 October, by rejecting the obesity pill lorcaserin, made by Arena Pharmaceuticals in San Diego, California. The agency has not approved a new obesity drug for more than a decade, and cited concerns about the drug's efficacy and side effects. The FDA's decision on another diet pill, Qnexa, developed by Vivus of Mountain View, California, is due on 28 October; an advisory panel voted against it in July. A third, Contrave, made by Orexigen in La Jolla, California, is up for FDA approval in December.

Avandia subpoena

In its third-quarter earnings report released on 21 October, drug giant GlaxoSmithKline (GSK) revealed that it is being subpoenaed by the US Department of Justice over the company's development and marketing practices for the diabetes drug Avandia (rosiglitazone). The company, headquartered in London, came under fire in July when a US Senate committee concluded that GSK had known about the drug's heart risks for more than a decade without reporting them to regulators. GSK denied the charge. Sales of Avandia are currently restricted in the United States and banned in Europe.

PEOPLE

Conflict of interest

Diána Bánáti, re-elected last week as chair of the management board of the European Food Safety Authority (EFSA), has resigned from the European board of directors of the International Life Sciences Institute, a non-governmental organization funded by food companies that seeks to coordinate and fund research and risk assessment. Her stepping down comes after controversy over alleged potential conflicts of interest (see *Nature* **467**, 647; 2010). The move was noted in an EFSA statement on 21 October.

Activists sentenced

Five British activists who tried to close down animal-testing firm Huntingdon Life Sciences near Cambridge, UK, by harassing and threatening anyone who did business with the company, were sentenced to between 15 months and 6 years in prison on 25 October. A sixth activist received a one-year suspended sentence. Seven other members of the same group, Stop Huntingdon Animal Cruelty, were sentenced in January 2009.
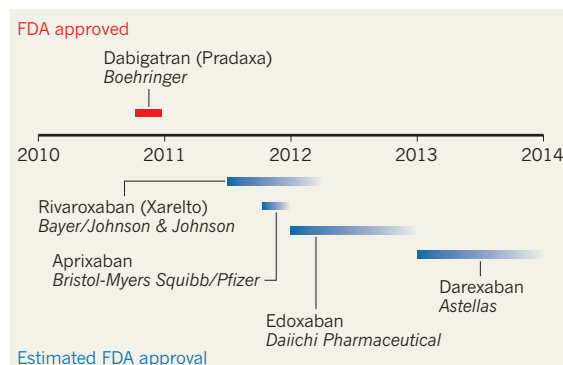

BUSINESS WATCH

Drug firms are racing to replace warfarin, a blood thinner in use since the 1950s. Many patients can't tolerate the drug, and its use requires regular blood tests. On 19 October, German firm Boehringer Ingelheim gained the US Food and Drug Administration's go-ahead to sell its drug dabigatran to some patients taking warfarin to prevent stroke. Other drug firms are not far behind (see chart). "It could be a very tightly fought battle," says Jonathan Angell, a market analyst at Datamonitor in London.

BLOOD-THINNING COMPETITION
Several firms hope to gain US approval for replacements to warfarin. Its market is worth some $400 million, but new drugs could earn billions of dollars, as they are costlier and applicable to more patients.

FDA approved
Dabigatran (Pradaxa)
Boehringer

2010 2011 2012 2013 2014

Rivaroxaban (Xarelto)
Bayer/Johnson & Johnson

Aprixaban
Bristol-Myers Squibb/Pfizer

Edoxaban
Daiichi Pharmaceutical

Darexaban
Astellas

Estimated FDA approval

NEWS IN FOCUS

T. DICKINSON

Comet Hartley 2 (right), shrouded by a glow of ionized gas, glides across a starry backdrop.

ASTRONOMY

Glimpsing a comet's heart

As comet Hartley 2 comes into close view, researchers are lining up with questions.

BY ADAM MANN

By now, the scenario is familiar: a distant light in the spacecraft's cameras becomes a fuzzy blob, which brightens and grows until the craft is suddenly plunging through an ionized fog. Enveloped in haze, the camera spies a dark, frozen lump — the elusive nucleus of a comet, one of the strangest and least understood bodies in the Solar System.

Since a battery of probes whizzed past comet Halley in 1986, the nuclei of four different comets have been successfully imaged and studied during fly-bys (see 'A gallery of surprises'). But rather than building up a simple and satisfying stereotype of what comets are like, these

encounters have revealed a surprisingly diverse array of features and processes. If all goes well, on 4 November, the cometary repertoire will grow by one more, when the NASA spacecraft EPOXI passes within 700 kilometres of comet Hartley 2 (see 'How to catch a comet').

"It seems like every time we go to a new comet, we discover new phenomena," says Lori Feaga, an astronomer at the University of Maryland in College Park, who is on the mission's science team.

In the annals of cometary exploration, EPOXI is already a hero. Formerly known as Deep Impact, in 2005 it flung a projectile at the nucleus of comet Tempel 1 and studied the plume of debris ejected by the impact. Since

then, it has been on course to Hartley 2 as part of the Deep Impact Extended Investigation (DIXI). During the five-year cruise it trained its camera on distant stars to search for signs of transiting exoplanets in a project called the Extrasolar Planet Observation and Characterization (EPOCh) investigation. A mash-up of acronyms gives the mission its current name.

Hartley 2 has already tantalized researchers with behaviour unlike anything seen at other comets, says principal investigator Michael A'Hearn, an astronomer at the University of Maryland who also led the Deep Impact encounter. Observing its target in September, the spacecraft discovered that Hartley 2's production of cyanogen — a byproduct ▶

of cyanides — increased fivefold over an eight-day period and then slowly returned to average. Such outgassing events on comets are usually violent and accompanied by dust, but this event was not, and the EPOXI team is still arguing about how to interpret the finding, A'Hearn says.

Anita Cochran, an astronomer at the University of Texas in Austin who studies Hartley 2 with ground-based telescopes, adds that the comet's nucleus, 1 kilometre in diameter, is putting out as much water vapour as Tempel 1, which has nearly ten times the surface area. She suspects that unlike larger comet nuclei with their isolated jets of gas and dust, Hartley 2's entire surface may seethe with outgassing. EPOXI scientists hope to learn why.

Such contrasts in appearance and behaviour challenge the notion that comets have a single, shared history. In the most general sense, they are understood to be accretions of frozen volatiles and rocky debris left over from the formation of the outer Solar System — fossils that preserve crucial information about the environment from which the outer planets emerged. With each close encounter, however, the picture becomes more complex.

The Stardust mission, for example, which collected material as it passed through the tail of comet Wild 2 in 2004 and brought the samples to Earth, found minerals that cou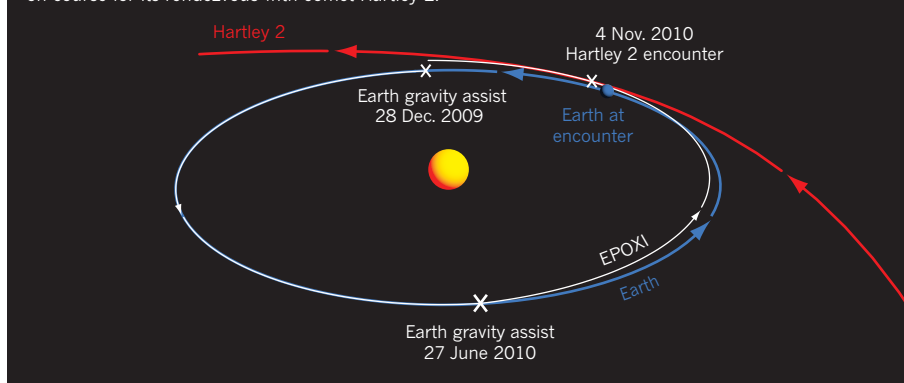ld only have been produced at high temperatures. This has led researchers to wonder if some comets were formed closer to the Sun than previously believed.

Deep Impact, for its part, identified 60 circular depressions on comet Tempel 1 that look like impact craters,

HOW TO CATCH A COMET
Two close encounters with Earth set NASA's EPOXI spacecraft on course for its rendezvous with comet Hartley 2.

Hartley 2

4 Nov. 2010
Hartley 2 encounter

Earth at encounter

Earth gravity assist
28 Dec. 2009

EPOXI

Earth

Earth gravity assist
27 June 2010

says A'Hearn. But the Sun's heat sublimates roughly half a metre of surface each time the comet completes an orbit, which should have quickly erased these marks. Another process must account for the depressions, according to A'Hearn.

The surface of Tempel 1 also showed what looked like cryo-volcanic flows, in which warmer, softer ice from the interior of the comet had apparently been extruded onto the frozen surface. "This seemed to indicate that some comet nuclei are active in their interiors," says Michael Belton, an emeritus astronomer at Kitt Peak National Observatory in Arizona. Belton and other researchers are developing theories to explain how cryo-volcanism could arise on such small, cold bodies.

Over the next five years, new missions are likely to add even more complexity to the cometary picture. In February 2011, the Stardust mission — rebranded NExT — is scheduled to revisit Tempel 1 to see how it looks five years after inspection by Deep Impact. Three

years later, Europe's Rosetta mission should reach comet Churyumov–Gerasimenko, and become the first spacecraft to orbit a comet nucleus and deposit a lander on its surface.

After that, comet science, which has flourished in recent years, could enter a lull without new missions to drive new discoveries. Such missions could be inherently more difficult and expensive than before— involving feats such as boring into a comet's nucleus — or take far longer to run. On the wish-list would be a journey to the comet reservoirs beyond Neptune's orbit to look at comets that are less altered from their pristine condition by successive passages near the Sun.

But comets remain a highly prized data source for many researchers. "NASA's stated goal is to explore the Solar System, which means you don't just go to the Moon and to Mars, you also explore unknown places," says David Jewitt, an astronomer at the University of California, Los Angeles. And comets, he says, "are really unknown places". ∎

A GALLERY OF SURPRISES
Four close encounters have yielded big differences among comets.
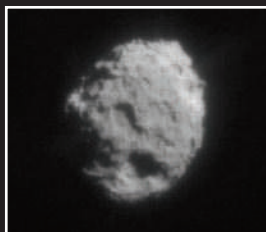
HALLEY (1986)

The first and largest (16 km) comet nucleus to be imaged, Halley showed bright jets and a nearly coal-black surface.

BORRELLY (2001)

Borrelly's patchy appearance hinted at variations in surface composition. Looking for ice, researchers found a warm, dry surface.

WILD 2 (2004)

Dust from the oddly pitted nucleus contained minerals that seemed to have formed nearer to the Sun than expected for a comet.

TEMPEL 1 (2005)

The best-imaged nucleus so far, Tempel 1 showed signs of cryo-volcanism and exposed ice. A probe found a surface fluffier than snow.

GLOBAL HEALTH

Verbal autopsy methods questioned

Controversy flares over malaria mortality levels in India.

BY DECLAN BUTLER

More than two-thirds of the world's population lives in countries that lack a reliable system for issuing medical death certificates, leaving the true scale and distribution of disease in serious doubt. The main tactic for filling that gap is verbal autopsy, which assigns a probable cause of death based on interviews with families about the deceased's symptoms.

But the reliability of the technique is under fresh scrutiny after a paper published in *The Lancet* last week[1] used verbal autopsy to calculate that 125,000–277,000 people in India die from malaria every year (see 'Malaria mortality'). That is an order of magnitude larger than the 30,000 deaths per year that the World Health Organization (WHO) estimates.

The *Lancet* paper used the most common form of verbal autopsy, in which physicians assign the cause of death. But statisticians argue that probabilistic computer models can do a better job than doctors. The WHO also argues that verbal autopsy can be poor at differentiating malaria from other diseases that cause fever symptoms, which include septicaemia, viral encephalitis and pneumonia. Although the WHO has accepted the use of verbal autopsy to monitor malaria deaths and other diseases, Christopher Dye, a senior WHO official, says the method can easily give misleading results.

Brian Greenwood, a malaria epidemiologist at the London School of Hygiene and Tropical Medicine, who performed some of the earliest verbal autopsies for malaria in Africa, says that malaria deaths in India are probably underestimated to some extent, but shares the WHO's concern about the "very poor" performance of the technique on fever symptoms.

Greenwood is also concerned that as physicians in the study were familiar with the Indian states that they reviewed case reports from, the survey had a built-in bias. As any medic in India probably knows the most malarious states, this could lead to "a temptation to ascribe febrile cases to malaria" in such states, says Greenwood.
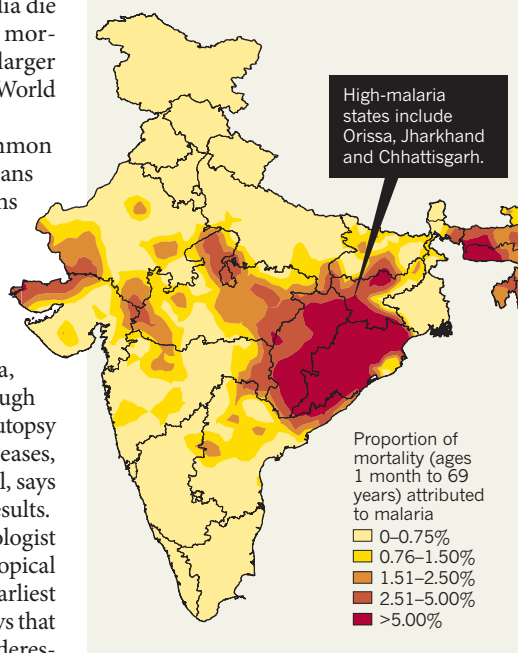
Prabhat Jha, an epidemiologist at the Centre for Global Health Research at the University of Toronto, Canada, and a co-author of the study, vigorously defends the results, arguing that physicians were given clear guidelines to carry out differential diagnosis to exclude malaria as the cause. The "total assignment of malaria deaths is not as biased as might be first believed", he says.

"We didn't blind as we thought it was important that coders knew where the case report came from," he adds. "It gave contextual information. If it smells like malaria, looks like

MALARIA MORTALITY

In 2001–03, malaria death rates in India were far higher than previously thought, according to a verbal-autopsy study.

High-malaria states include Orissa, Jharkhand and Chhattisgarh.

Proportion of mortality (ages 1 month to 69 years) attributed to malaria
- 0–0.75%
- 0.76–1.50%
- 1.51–2.50%
- 2.51–5.00%
- >5.00%

malaria, and you see it in malarious regions then it probably is malaria."

But Gary King, a statistician at Harvard University in Cambridge, Massachusetts, notes that the different pairs of physicians that looked at each case in the *Lancet* paper often disagreed on the cause of death. "The error rates between the experts account for half the malaria deaths estimated," he says.

Bob Snow, a malaria epidemiologist at the Kenya Medical Research Institute–Wellcome Trust Research Programme in Nairobi, says that whatever the limitations of the study, its estimates are "closer to the truth than the WHO figures", and that its findings are consistent with the spatial and temporal epidemiology of malaria in India. Snow notes that the paper is in line with his own team's findings that the WHO has underestimated the clinical incidence of malaria in India by a similar order of magnitude[2].

THE NEEDS OF THE MANY

Verbal autopsy is increasingly being questioned by statisticians, says Edward Fottrell, an epidemiologist at Umeå University in Sweden. Until now, verbal autopsy has been dominated by physicians, whose clinical background means that they tend to believe that diagnosing individual cases is key for accuracy, he says.

But the ultimate goal of verbal autopsy is not to make clinical diagnoses of individual cases, Fottrell points out. It is to estimate the distribution of causes of deaths, known as cause-specific mortality fractions (CSMFs), which are crucial to setting health-system and research priorities, and to monitoring the effectiveness of disease-control measures.

Pigeonholing cases into a single, accurate cause of death can amplify the errors in the CSMFs, says King. A better approach, he says, is to calculate the probabilities that various disease symptoms are associated with a death, and then aggregate those probabilities across an entire set of cases[3].

Studies show that these probabilistic computer models can give CSMFs as good as or better than physician review, but are far faster and cheaper[4]. They also overcome the issue of physician subjectivity, providing a standardized method that makes results more comparable between different studies and countries.

Many researchers are reluctant to embrace verbal-autopsy models that dispense with physician review, but attitudes may be changing. The Swedish International Development Cooperation Agency, based in Stockholm, recently recommended that the international INDEPTH surveillance network, which records births, deaths and disease within large population cohorts in 17 African and Asian countries, adopts a probabilistic verbal-autopsy model. Fottrell predicts that computer models will eventually prevail over physician review.

The ultimate goal, however, is to ensure that verbal autopsy is no longer needed, says Dye, and the WHO is helping all countries to eventually implement the gold standard of a systematic medical death certification. "That is the end point that the WHO is working towards." ∎

1. Dhingra, N. *et al. Lancet* doi:10.1016/S0140-6736(10)60831-8 (2010).
2. Hay, S. I. *et al. PLoS Medicine* **7**, e1000290 (2010).
3. King, G., Lu, Y. & Shibuya, K. *Popul. Health. Metr.* **8**, 19 (2010).
4. Fottrell, E. & Byass, P. *Epidemiol. Rev.* **32**, 38–55 (2010).

RESEARCH FUNDING

UK scientists celebrate budget reprieve

Core science funding has escaped cuts, but capital budgets will feel the squeeze.

BY GEOFF BRUMFIEL

An unexpected bouquet of white lilies and roses greeted David Willetts, Britain's minister for science, when he arrived at a press conference on 20 October to announce the government's plans for research spending over the next four years.

In better times, he might have been met with a barrage of rotten fruit. The research base will continue to be funded at its current level, £4.6 billion (US$7.2 billion), for the four-year review period — which amounts to an effective cut of 10% if inflation projections are factored in. In addition, an essential funding stream for large projects will probably be substantially cut, along with research in many government departments.

But these are not better times. Faced with a record deficit of £109 billion, the British government is slashing expenditure by an average of 19% across its departments. In the face of such austerity, Willetts called the science budget a "fantastic deal", and many agreed. "I'm genuinely relieved," says William Cullerne Bown, founder of the science-policy newsletter *Research Fortnight*, who presented Willetts with the flowers. John Beddington, the government's chief scientific adviser, says that officials such as George Osborne, the Chancellor of the Exchequer, were won over by arguments from high-profile scientists and industrialists that cuts could hinder long-term growth of the British economy.

The £4.6-billion sum includes funding for the nation's research councils, which dole out grants to scientists, and money for 'quality related' research funds, which universities can prioritize as they choose. Money for health research — channelled through the Department of Health, and the Medical Research Council (MRC) — will remain flat in real terms (once inflation is factored in),

amounting to a modest increase in cash terms. Other research councils will have to bear a greater burden of cuts to compensate for the MRC's good fortune. All funding has been assured for the four-year period, according to Willetts.

The budget also provides £220 million for the research councils' highest future priority — a medical research centre to be located in the heart of London. Documents obtained by *Nature* under freedom of information legislation show that the councils deemed the UK Centre for Medical Research & Innovation such a high priority that they declined to even rank it against other projects when submitting budget documents earlier this year. An upgrade to the Diamond synchrotron in Oxfordshire is also assured. "The outcome is better than most of us had hoped for," says Martin Rees, president of the Royal Society, Britain's national science academy.

But money for infrastructure and subscriptions to large international projects is not protected, according to Willetts. The Department for Business, Innovation and Skills, which funds the councils, will see its overall 'capital' budget fall by 44% to £1 billion in 2014–15 (see 'Capital crash').

> *"The outcome is better than most of us had hoped for."*

That money pays for everything from radio telescopes to Antarctic research stations. In particular, the cuts will hit the Science & Technology Facilities Council (STFC), which funds particle physics and astronomy. The council, which has struggled financially for years, has been told to prepare for its capital funding to fall by a third, according to documents seen by *Nature*. That could jeopardize Britain's participation in organizations such as the European Southern Observatory.

CAPITAL CRASH
Infrastructure funds at the Department for Business, Innovation & Skills, about half of which go to research, will fall sharply in the coming years.
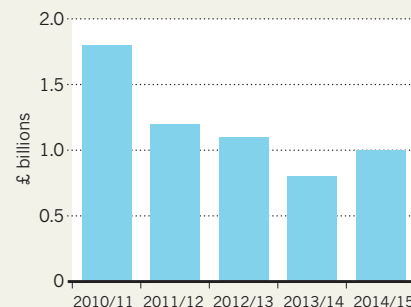
Research funding in government departments will also be under pressure. The annual £650-million basic-research budget of the Ministry of Defence will probably face a "modest" cut, says Willets. The Department for Environment, Food and Rural Affairs, which conducts animal health and environmental research, will face "substantial but manageable" cuts to its £95-million annual core research budget, according to Chris Gaskell, who heads the department's independent scientific advisory council. Beddington says that he will be consulted before any departmental cuts are made final. "It doesn't mean I can veto them, but it does mean that it will be discussed," he says.

The final details of what is cut, and how, will emerge in the weeks and months to come (see *Nature* **467,** 894; 2010), but for now, the mood is buoyant. After handing his flowers to an aide, Willetts turned to the assembled reporters and policy-makers with a broad smile. "We'll have the hugs and kisses later on," he joked. ∎

SEE WORLD VIEW P.1007

MORE ONLINE

LATEST NEWS
- NASA charged with leading asteroid response go.nature.com/hedzcu

MORE STORIES
- Rise in space tourism could accelerate climate change go.nature.com/Ifigec
- Alternative yardstick to measure the Universe go.nature.com/fmxrog
- Ancient chimp virus brought 'back to life' go.nature.com/qhbqOK

IN CONVERSATION
Former US science adviser Nina Fedoroff explains why people can achieve more than governments. go.nature.com/rxxhdr

D. VAN RAVENSWAAY/SPL

D. MILLS/NYT/REDUX/EYEVINE

REUTERS/ST-FELIX EVENS

The devastation created by Haiti's magnitude-7 earthquake left 1.3 million survivors homeless.

EARTH SCIENCE

Quake threat looms over Haiti

Tectonic strain remains in key fault line, researchers find.

BY QUIRIN SCHIERMEIER

The half-minute of tremors that shook Haiti in January left death and destruction — and lingering questions about when and where another such quake might strike. Some 230,000 people died in the magnitude-7.0 quake, more than twice as many as in any recorded earthquake of similar strength. As the disaster drew aid workers from around the globe, scientists also flocked to the impoverished country to try to understand the quake.

What they found was unexpected. After ten months of intense field research, geologists are questioning conventional wisdom about what happened to Earth's crust during the fateful 30 seconds that set back Haiti's development by years. The research, summarized in a package of papers in the November issue of *Nature Geoscience*, has two common conclusions: the Haitian earthquake was more complex than initially believed, and may not have fully released the tectonic strain that had accumulated in the

region. If so, Haiti is at serious risk of similar devastation in the future.

"The 12 January earthquake only unloaded a fraction of the seismic energy that has built up over time in Haiti," says Eric Calais, a geophysicist at Purdue University in West Lafayette, Indiana, and science adviser for the United Nations Development Programme in Haiti. "Other earthquakes are therefore inevitable."

The Haiti quake occurred in a Caribbean fault system called the Enriquillo–Plantain Garden, at the interface of the Caribbean and North American plates, where seismic strain gradually accumulates as the two plates slide past each other (see 'Anatomy of a quake'). Strong earthquakes originating from this fault have twice destroyed Port-au-Prince, in 1751 and 1770. Using computer models alongside satellite and field observations, Calais and other scientists have tried to establish which parts of the fault system ruptured this time around, and in which direction.

The results suggest that the quake may not

have originated from the main fault in the system, as geologists had initially assumed. For example, there is a puzzling absence of the geological evidence normally left by tectonic slips that rupture the surface. A team led by Carol Prentice of the US Geological Survey (USGS) in Menlo Park, California, spent months searching the land along the plate boundary fault south of Port-au-Prince for such traces. Although they found stream channels that had been wrenched sideways during historic quakes, they failed to find any fresh signs of surface rupture around the main fault[1].

"This is pretty bizarre," says Roger Bilham, a geologist at the University of Colorado, Boulder, who was not involved in the recent studies. "It might mean that the main fault is a geological fossil. But more likely its surface part has been clamped shut by a complex sequence of nearby slips in January. If so, another strong quake could happen any time soon right above the January epicentre."

The findings also mean that the January quake must have been triggered along another fault. To pinpoint it, two teams of scientists have created different fault models based on ground deformation, seismic waves recorded at the time, and the little that is known about local geology. Unsurprisingly, given the uncertainties in the data, the models differ considerably.

⟳ **NATURE.COM**
For more, see
Nature Geoscience:
www.nature.com/ngeo

Calais' team says that the quake occurred on a previously unknown subsidiary fault in the Enriquillo–Plantain Garden.

Dubbed Léogâne, after a nearby town, it lies to the north of and parallel to the main fault[2].

The second team, led by Gavin Hayes, a seismologist with the USGS in Golden, Colorado, reckons that the quake involved at least three faults, which mutually triggered each others' slipping. The slip started on either the main Enriquillo fault or the Léogâne subsidiary fault, they conclude[3].

To assess the hazard of future quakes in the region, scientists need to know how much additional seismic stress was transferred to nearby faults by January's disaster. But that assessment would vary depending on the model used — an uncertainty that offers little comfort for planners and engineers in Haiti, or for the 1.3 million survivors living in camps after their homes were destroyed. As *Nature* went to press, those people were facing the growing threat of a rapidly spreading cholera outbreak.

The January quake also had unexpected effects at the surface. Scientists led by Susan Hough of the USGS in Pasadena, California, have found that the strongest ground motion did not occur in the soft sedimentary rock that underlies most of Port-au-Prince, as would be expected. Instead, the greatest movement was seen in a foothill ridge south of the capital, where the ground consists of relatively solid rock[4]. The team believes that seismic waves were amplified by local geological conditions

ANATOMY OF A QUAKE
Researchers led by geophysicist Eric Calais say that Haiti's recent earthquake originated from the previously unknown Léogâne Fault.

and topographic features such as valleys and hills.

"What we know now hasn't brought us any closer to understanding Haiti's seismic future," says Bilham. "As things stand, we can only recommend engineers rebuild Port-au-Prince as safely as money allows." An array of seismic instruments installed across Haiti since the quake may soon provide some of the missing information about the fault's origin, and the amount of strain remaining in the system, he adds. The array is recording frequent tiny quakes, of magnitudes 1–2, which will help scientists to map the region's subsurface geometry and improve their models.

"We know enough already to recommend

proactive measures to adapt the country to earthquake hazard and, eventually, reduce economic losses and save lives," says Calais. "But research must continue to better characterize seismic hazard. A dedicated effort is key to identifying all potential sources of earthquakes and producing the hazard maps that are badly needed for planning and engineering purposes." ∎

1. Prentice, C. *et al. Nature Geosci.* doi:10.1038/ngeo991 (2010).
2. Calais, E. *et al. Nature Geosci.* doi:10.1038/ngeo992 (2010).
3. Hayes, G. P. *et al. Nature Geosci.* doi:10.1038/ngeo977 (2010).
4. Hough, S. E. *et al. Nature Geosci.* doi:10.1038/ngeo988 (2010).

PUBLISHING

Mystery fraud accusations

Stem-cell researchers targeted by e-mails from unidentified group.

BY HEIDI LEDFORD

It's a researcher's worst nightmare: an unexpected allegation of scientific misconduct broadcast to colleagues and journalists without any clue as to where the accusation is coming from or how to respond to it. That's what happened twice last week, when a group calling itself 'Stem Cell Watch' sent e-mails claiming evidence of fraud in recent publications from prominent stem-cell researchers. "We are continuing to point out suspicious results and duplications reported by scientists in the stem-cell field," the group wrote.

There is no indication that any of its accusations are correct, but the group has rattled a rapidly moving field that is accustomed to controversy, causing researchers to fear for their credibility and forcing journal editors to re-examine published work. (The International Cellular Medicine Society also runs a website called Stem Cell Watch, which has no association with the e-mail group.)

At least three research teams have found themselves in the cross hairs of Stem Cell Watch, and the group says it is considering action against others. But its behaviour is raising the hackles of scientists, who believe the alerts are smearing reputations without cause. "I find this kind of activity unhelpful and defamatory," says Doug Melton, co-director of the Harvard Stem Cell Institute in Cambridge, Massachusetts.

Critics argue that Stem Cell Watch is not following scientific etiquette, which says that concerns should be addressed directly and openly to the authors of a paper. Melton says he received a message, addressed only to him, from the group earlier this year. The e-mail accused another stem-cell researcher of misconduct, but because it was anonymous, Melton simply deleted it.

Stem Cell Watch provides little information about its members. They claim to be students majoring in biology who discuss papers taught in class. Their aim, they say, is to alert professionals to problems they find in the literature, to ensure that they are handled seriously.

One of Stem Cell Watch's missives last week stated that images of the same cells had been used more than once, but with different coloration, in a 2009 paper in the *Proceedings of the National Academy of Sciences* (S. Friling *et al. Proc. Natl Acad. Sci. USA* **106,** 7613–7618; 2009). Indeed they were the same cells, retort the corresponding authors,
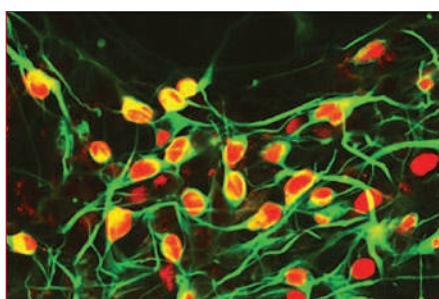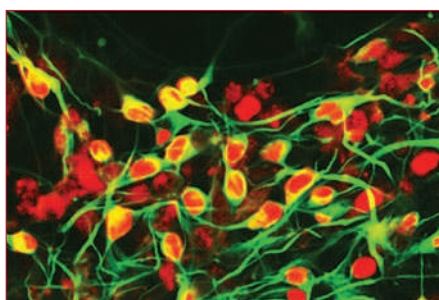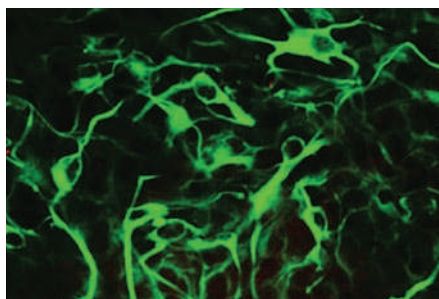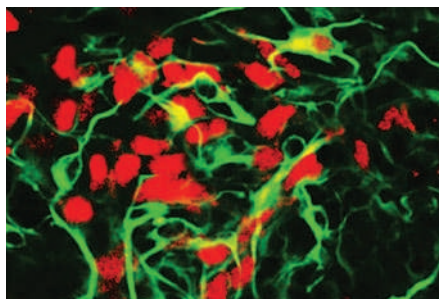
Anonymous tipsters forced researchers to defend the use of the same cells in multiple images.

S. FRILING *ET AL. PNAS* **106,** 7613–7618; 2009

Johan Ericson and Thomas Perlmann at the Karolinska Institute in Sweden, but the images were appropriate because multiple proteins in the cells had been labelled with differently coloured fluorescent tags. "We appreciate any opportunity to respond to critique or concerns raised about our work," Perlmann and Ericson said in a written statement. "However, we regret that these serious accusations were made anonymously, as we strongly believe in the concept of an open and transparent communication about suspected errors in published data." A spokeswoman at the *Proceedings of the National Academy of Sciences* says that the journal is obliged to investigate the group's claims as a matter of policy.

In another e-mail, Stem Cell Watch attacked a 2009 paper in *Nature* in which Konrad Hochedlinger at the Harvard Stem Cell Institute and his colleagues reported a new link between the generation of induced pluripotent cells and cancer (J. Utikal *et al. Nature* **460,** 1145–1148; 2009). The group says it decided to take action after "several conceptual flaws" led them to evaluate the paper's images more closely. Their e-mail states that the images in one figure "appear weird", and that the same embryo is probably depicted in the figure's control and experimental panels. The anonymous accusers also asserted that the fluorescence staining in the experimental panel "appears very artificial to the experienced eye" and may have been "introduced by fraudulent computer photo manipulation or other means". The message concluded with a call for *Nature* to investigate the matter.

"We wouldn't encourage anonymous accusations, least of all those broadcast indiscriminately," says Philip Campbell, *Nature*'s editor-in-chief. "But there have been occasions where anonymous whistle-blowing has revealed fraudulent papers, so we will at least consider such accusations."

Hochedlinger was caught by surprise by the nature of the accusation. "I have never received e-mails like this before and, to be honest, I find it quite upsetting," he says. He has reviewed the original images and says the allegations are entirely unfounded. He has submitted the originals to *Nature* to assist with any review.

As before, the accusations seem unlikely to be valid. The *Nature* paper was one of several published simultaneously by different research groups reporting similar results. Although *Nature* has not commented on the specific allegations, five stem-cell researchers contacted by a *Nature* reporter say they saw no evidence of fraud in either the original images or the figure as presented in publication. Whoever the group is, says Robin Lovell-Badge at the National Institute for Medical Research in London, "it seems they do not have that much experience looking at mouse embryos".

Lovell-Badge adds that he finds the incident worrying. "Although we don't want fraudulent work to be published," he says, "this group does not seem to have the skill or knowledge to make a fair assessment." ∎

ENVIRONMENT

Mountaintop mining plans close to defeat

Environmental review details 'unacceptable' impacts.

BY NATASHA GILBERT

The rising tide of scientific evidence — and public protest — against mountaintop mining looks set to claim its first major victory. By the end of this year, the US Environmental Protection Agency (EPA) is expected to revoke a permit allowing mining company Arch Coal to extract coal from the Appalachian Mountains in West Virginia. This would be the first time a permit for the controversial mining practice, long suspected of causing environmental damage, has been vetoed by the agency.

A scientific review (see go.nature.com/hsuhrt) carried out by the EPA and published on 15 October concluded that the project, Spruce 1, would have "unacceptable" effects on water quality and wildlife, and recommended its permit be revoked. Carol Raulston, a spokeswoman for the National Mining Association (NMA), based in Washington DC, told *Nature*: "The NMA has no reason to believe the EPA will not follow the recommendations in its final determination on the Spruce permit."

The move is likely to set the tone for decisions on other mining projects. More than 100 surface-mining permits are pending approval with the Army Corps of Engineers, which is responsible for investigating, developing and maintaining the nation's water and related environmental resources. The corps issued approval for the Spruce 1 project in 2007 to Mingo Logan, a subsidiary of Arch Coal. But the EPA can revoke a permit if it feels that environmental concerns have not been fully addressed.

"If the EPA proceeds with its unlawful veto of the Spruce permit, as it appears determined to do, every business in the nation would be put on notice that any lawfully issued permit can be revoked at any time according to the whims of the federal government," says Kim Link, a spokeswoman for Arch Coal.
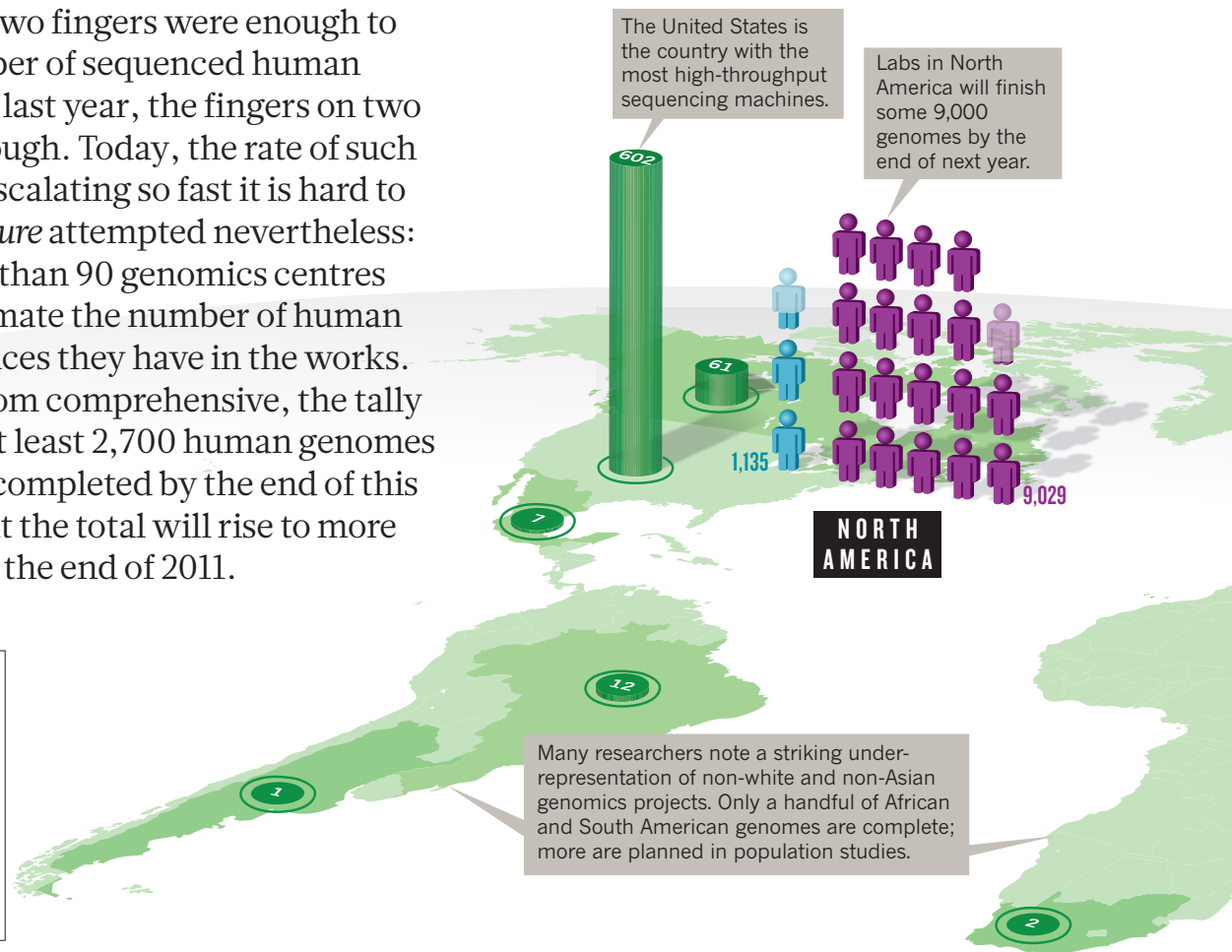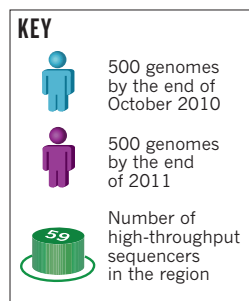
Mountaintop mining exposes seams of coal near mountain peaks by stripping away forests and breaking up rock with explosives. The debris is often dumped in the valleys below. The EPA review says that Spruce 1 would increase the electrical conductivity of stream water (a measure of its ionic concentration) to unacceptably high levels, harming aquatic wildlife.

The NMA says that the EPA's use of electrical conductivity as a proxy for water pollution is "faulty science". "Conductivity is but one metric of water quality and is not recognized by hydrologists as satisfactory when used as the chief or only metric," says Luke Popovich, a spokesman for the NMA. However, research has shown a strong correlation between increased levels of conductivity and harm to aquatic macro-invertebrates (see *Nature* **466,** 806; 2010).

Arch Coal had already filed a lawsuit in April challenging the EPA's authority to veto permits. The company now plans to submit a rebuttal to the review by 5 November. ∎

Genomes by the thousand

Ten years ago, two fingers were enough to count the number of sequenced human genomes. Until last year, the fingers on two hands were enough. Today, the rate of such sequencing is escalating so fast it is hard to keep track. *Nature* attempted nevertheless: we asked more than 90 genomics centres and labs to estimate the number of human genome sequences they have in the works. Although far from comprehensive, the tally indicates that at least 2,700 human genomes will have been completed by the end of this month, and that the total will rise to more than 30,000 by the end of 2011.

The United States is the country with the most high-throughput sequencing machines.

Labs in North America will finish some 9,000 genomes by the end of next year.

602

61

1,135

7

9,029

NORTH AMERICA

12

1

2

Many researchers note a striking under-representation of non-white and non-Asian genomics projects. Only a handful of African and South American genomes are complete; more are planned in population studies.

KEY

- 500 genomes by the end of October 2010
- 500 genomes by the end of 2011
- 59 — Number of high-throughput sequencers in the region

Why scientists want tens of thousands of genomes — and more

To understand populations

Comparing lots of genomes lets researchers identify points at which one genome differs from the next. Costs may be falling, but sequencing and data analysis are still pricey. So most researchers face a trade-off between the number of subjects and the accuracy in the sequences they can afford. For projects examining how populations commonly differ, sequencing a large number of individuals at relatively low accuracy or 'depth of coverage' is enough. About 900 genomes sequenced so far by the 1000 Genomes Project have been read three times on average.

Cost still limits the accuracy with which genomes are sequenced.

High coverage (30× and above) 974

Low & medium coverage (15× and below) 1,756

GENOME ACCURACY
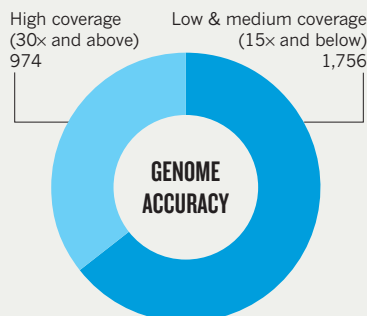
To understand disease

Researchers trying to uncover rare disease-linked mutations — perhaps limited to just one family or an individual — need precision, typically sequencing each genome 30 times on average. Cancer genomes, many sequenced under the auspices of large collaborations, account for a sizeable chunk of high-coverage genome sequences completed to date. Projects scrutinizing people with diabetes, Crohn's disease and other disorders are starting to emerge. Analysing all the genome data is a huge challenge, as is turning genetic discoveries into clinical benefits.

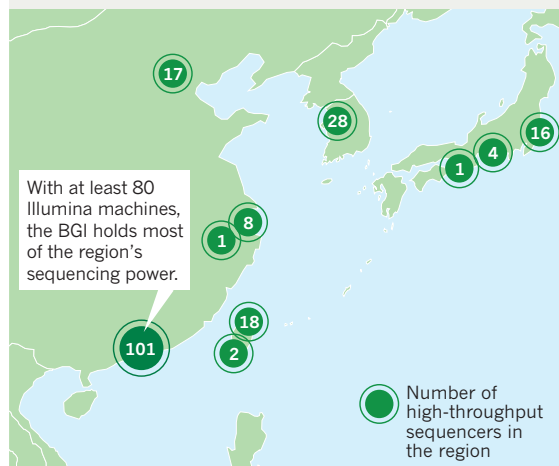GRAPHICS BY NIK SPENCER; DATA COMPILED BY ALLA KATSNELSON

POWER TO THE PEOPLE

The latest sequencing technology is no longer concentrated at a few major centres. In Britain, the Wellcome Trust Sanger Institute in Hinxton houses 38 of the country's high-throughput sequencers, and the rest are scattered over an additional 32 sites. Falling costs mean that a human genome is within the reach of individual labs.
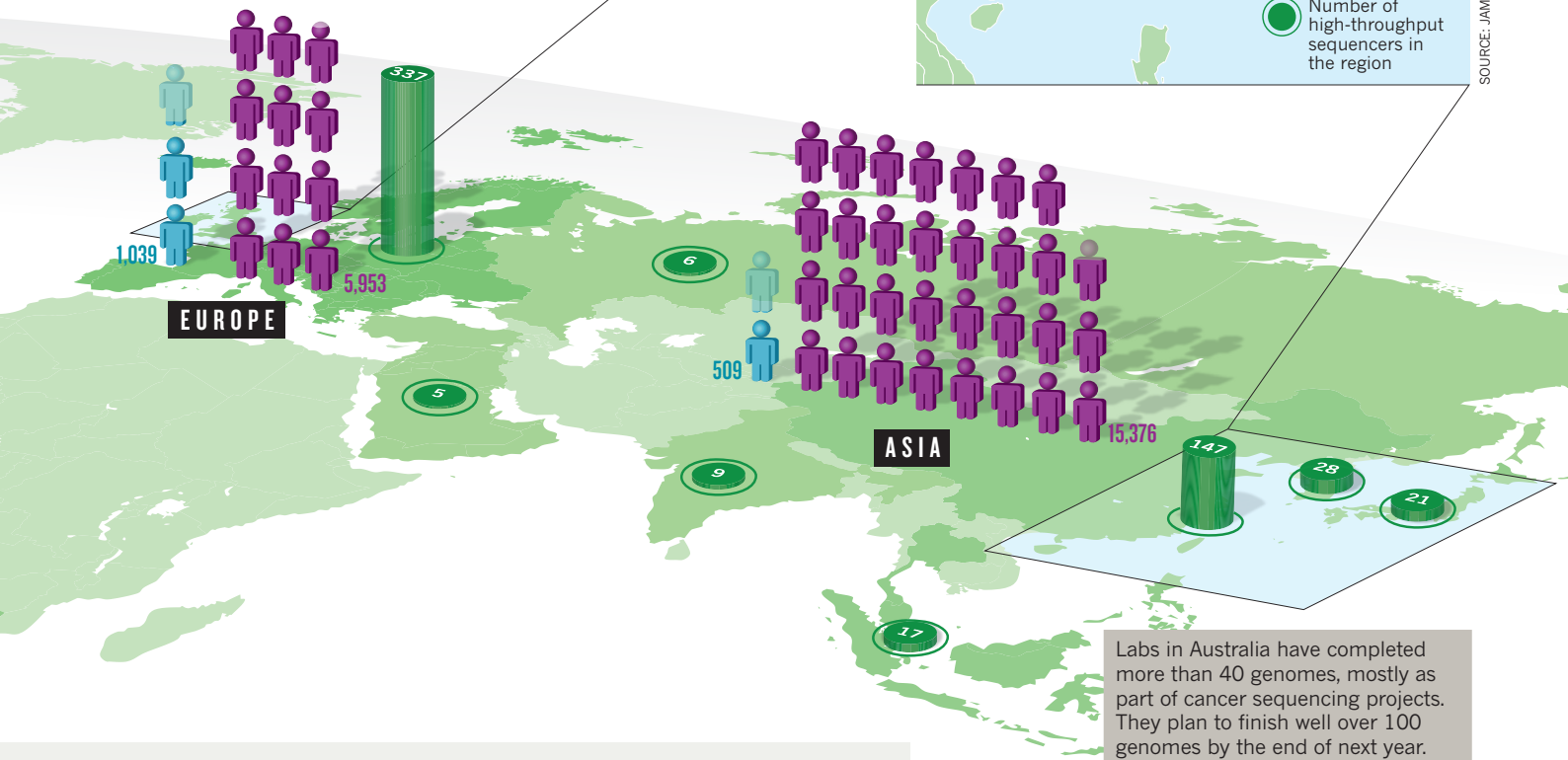
UK machines will help to sequence 2,500 genomes as part of the 1000 Genomes Project.
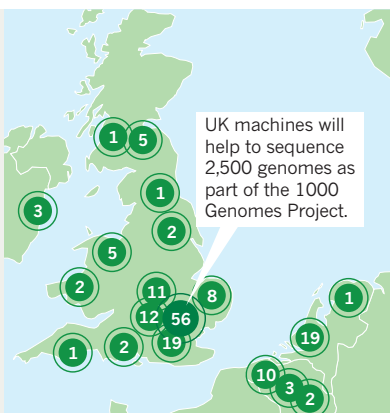
THE RISE OF GENOME FACTORIES

Individual labs may still find it cheaper and easier to outsource a human genome to a power-house 'sequencing service provider'. The BGI in Shenzen, which has global expansion plans, predicts that its machines will have completed some 10,000 to 20,000 human genomes by the end of 2011.

With at least 80 Illumina machines, the BGI holds most of the region's sequencing power.

Number of high-throughput sequencers in the region

SOURCE: JAMES HADFIELD & NICK LOMAN

EUROPE
1,039
5,953
337

ASIA
509
15,376
6
5
9
17
147
28
21
35
3

Labs in Australia have completed more than 40 genomes, mostly as part of cancer sequencing projects. They plan to finish well over 100 genomes by the end of next year.

Disease-specific projects make up more than half of the complete genomes.

Personal 44
Other/unknown 55
Cancer 485
Other diseases 916
Population 1,230

TYPE OF PROJECT
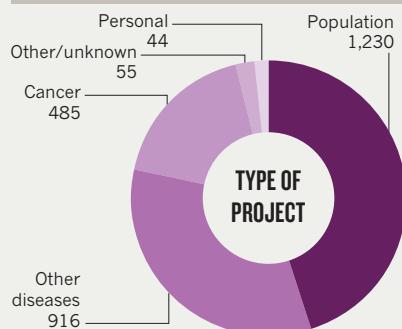
To understand individuals

The rate at which human genomes are being sequenced — at least in mega-projects — will probably slow once researchers have extracted most of the common variation shared by populations and diseases. But individuals are genetically unique. If the cost of a genome sequence becomes trivial and the benefits of knowing one increase (through gene-tailored medicine), then personal genome sequencing will continue to push the genome count up and up.

NATURE.COM
For more on human genomes see:
www.nature.com/humangenome

METHODS: Our survey focused on large, academic projects rather than individual labs; we included complete genome sequences, both high- and low-coverage, and excluded partial (exome) sequences. The list excludes all biotechnology and pharmaceutical companies and most sequencing service providers, which do not disclose their work. The sequencer locations, based on a user-generated map at go.nature.com/b74acy, includes some 60–70% of all machines.

A NASA technician prepares six of the James Webb Space Telescope's mirror segments for cryogenic testing.

THE TELESCOPE THAT ATE ASTRONOMY

NASA's next-generation space observatory promises to open new windows on the Universe — but its cost could close many more.

BY LEE BILLINGS

It has to work — for astronomers, there is no plan B. NASA's James Webb Space Telescope (JWST), scheduled to launch in 2014, is the successor to the Hubble Space Telescope and the key to almost every big question that astronomers hope to answer in the coming decades. Its promised ability to peer back through space and time to the formation of the first galaxies made it the top priority in the 2001 astronomy and astrophysics decadal survey, one of a series of authoritative, ten-year plans drafted by the US astronomy community. And now, the stakes are even higher. Without the JWST, the bulk of the science goals listed in the 2010 decadal survey, released this August, will be unattainable.

"We took it as a given that the JWST would be launched and would be a big success," says Michael Turner, a cosmologist at the University of Chicago, Illinois, and a member of the committee for the past two decadal surveys. "Things are built around it."

Hence the astronomers' anxiety: the risks are also astronomical. The JWST's 6.5-metre primary mirror, nearly three times the diameter of Hubble's, will be the largest ever launched into space. The telescope will rely on a host of untried technologies, ranging from its sensitive light-detecting instrumentation to the cooling system that will keep the huge spacecraft below 50 kelvin. And it will have to operate perfectly on the first try, some 1.5 million kilometres from Earth — four times farther than the Moon and beyond the reach of any repair mission. If

the JWST — named after the administrator who guided NASA through the development of the Apollo missions — fails, the progress of astronomy could be set back by a generation.

And yet, as critical as it is for them, astronomers' feelings about the JWST are mixed. To support a price tag that now stands at roughly US$5 billion, the JWST has devoured resources meant for other major projects, none of which can begin serious development until the binge is over. Missions such as the Wide-Field Infrared Survey Telescope, designed to study the Universe's dark energy and designated the top-priority space-astronomy project in the most recent decadal survey, will have to wait until after the JWST has launched. "Until then, we're not projecting being able to afford large investments" in new missions, says Jon Morse, director of NASA's astrophysics division. And all the space telescopes currently operated by NASA and the European Space Agency will reach the end of their planned lifetimes in the next few years.

Worse, the JWST's costs keep growing. In 2009, NASA required an extra $95 million to cover cost overruns on the telescope. In 2010 it needed a further $20 million. And for 2011 it has requested another $60 million — even as rumours are swirling that still more cash infusions will be required (see 'Cost curve').

Senator Barbara Mikulski (Democrat, Maryland), chairwoman of the government subcommittee that oversees NASA's budget, responded to these requests in June by calling for an independent panel to investigate the causes of the JWST's spiralling cost and delays, and to find a way

to bring them to resolution. "Building the JWST is an awesome technical challenge," Mikulski says. "But we're not in the business of cost overruns."

John Casani, chairman of Mikulski's investigative panel and a former project manager for NASA's Voyager, Galileo and Cassini missions, emphasizes that the panel is making suggestions, not decisions. Those will be up to NASA, which is expected to announce a budgetary plan incorporating the panel's suggestions on 2 November. But in considering potential solutions for the JWST's woes, Casani says that "everything will be on the table" — including, conceivably, scrapping instruments or otherwise downgrading the programme.

THE GOLDIN OPPORTUNITY

The first concept for a Hubble replacement emerged in 1989, when Hubble was still a year away from launch. Astronomers already knew that its vision would not quite reach back to the 'cosmic dawn', 500 million years after the Big Bang, when the first stars and galaxies formed. So a next-generation space telescope that could fill the gap seemed like the logical next step.

In 1993, NASA asked a committee of astronomers, chaired by Alan Dressler of the Carnegie Observatories in Pasadena, California, to define what such a telescope would need. The new telescope's mirror would have to be big to gather the dim light of those first galaxies. So the committee recommended that the primary mirror be at least 4 metres across.

The telescope would also have to be cryogenically cold, because at any temperature higher than 50 kelvin, infrared heat radiation from the telescope itself would wash out the faint photons that the astronomers were looking for. "That was the science that propelled the whole thing," says Dressler.

Finally, it would have to operate far from Earth. At infrared wavelengths, this planet glows like a light bulb. So the committee recommended that the telescope be placed 1.5 million kilometres outside Earth's orbit, at the second Lagrangian point (L_2), where the combined gravitational pull of the Sun and Earth creates a region of stability. Any spacecraft at L_2 will also lie in the shadow cast by Earth, making it easier to keep cool (see 'The James Webb Space Telescope').

In December 1995, Dressler briefed NASA's then administrator, Daniel Goldin, on the recommendations. Goldin was intrigued. He was shaking up NASA's science programmes, pushing a 'faster, better, cheaper' strategy to deliver more capable and inspiring missions at lower costs. Taking his cues from Silicon Valley and aerospace 'skunkworks' projects — small, highly autonomous ventures pursuing innovation within larger organizations — Goldin was pushing for miniaturization of bulky electronics, more off-the-shelf components, lower organizational overheads, and a continuous expansion of the technological boundaries with each mission. Dressler's proposal seemed like a perfect opportunity to test that approach.

Instead of a 4-metre telescope, Goldin asked, why not try one with a primary mirror 6–8 metres in diameter? Some of the technology was in hand: NASA was developing the cryogenic infrared Spitzer Space Telescope with a 0.85-metre mirror made of beryllium, a metal that needs special handling — it corrodes skin at a touch — but is lightweight and keeps its shape through extreme temperature changes. That and other innovations could give the JWST a mega-mirror while reducing costs. As Goldin put it in a speech: "Let's throw away glass. Glass is for the ground."

Some astronomers were dubious about initial cost estimates for the ambitious mission, which ranged from $500 million to $1 billion. But in the beginning, Goldin's methods seemed to deliver: the

first missions using the approach were wildly successful. Among them were 1997's landmark Mars Pathfinder mission and its accompanying rover, Sojourner, and the 1998 Lunar Prospector mission that found evidence of water ice on the Moon. But they were followed in 1999 by the disastrous losses of the Wide-Field Infrared Explorer telescope and two planetary missions, the Mars Climate Orbiter and the Mars Polar Lander. This string of failures tarnished the agency's reputation, and reminded everyone that 'faster, better, cheaper' was also riskier. By the end of Goldin's tenure in 2001, NASA had already begun shifting back to its traditional, risk-averse and far more expensive strategy of exhaustive testing and extensive oversight.

That shift would send the cost of the JWST soaring past the billion-dollar mark. The mirror diameter would be cut from 8 metres to 6.5 metres to help reduce costs. But in the meantime, as NASA carried out the many engineering trade-off studies and scientific working groups required to solidify the telescope's design, a more insidious factor came into play: scientists started to pile on complexity.

It happens with almost every major mission, says Peter Stockman, former head of the JWST mission office at the Space Telescope Science Institute in Baltimore, Maryland. "Everyone fears it will be the last opportunity in their scientific lifetime." And there seemed little reason for restraint: in the 1990s, when the bulk of the design work was done, NASA's astrophysics budget was projected to keep growing by a few per cent a year.
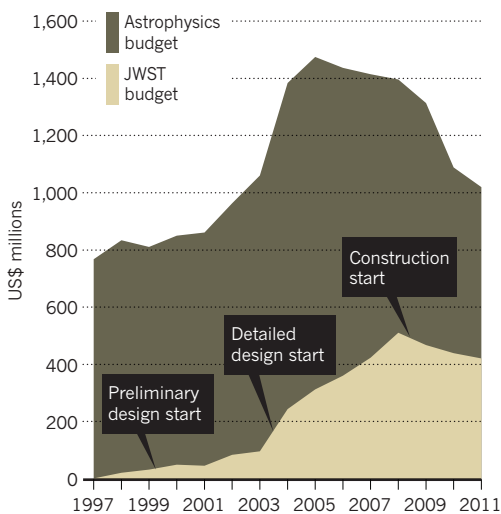
STRETCHED CAPABILITIES

With each iteration, the JWST's science objectives swelled. The core instrument package came to include a large-field-of-view near-infrared camera (NIRCam) and a multi-object near-infrared spectrograph (NIRSpec), primarily for investigating the earliest stars and galaxies; a general-purpose mid-infrared camera and spectrograph for observing dust-shrouded objects in the Milky Way; and a fine guidance sensor and tunable-filter imager to support the other three.

These expanded capabilities would have to be supported by expensive and largely unproven technologies. The instruments needed extra-large, ultra-stable infrared detectors. A five-layered membranous sunshield would have to be folded around the spacecraft before launch, then deployed in space to allow the telescope to cool to cryogenic temperatures. Unfurled, each layer would be about the same area as a tennis court. The primary mirror, too large to fit into any existing rocket fairing, would have to be assembled in 18 hexagonal, adjustable segments that would also unfold in orbit. Each segment would be painstakingly chiselled from beryllium, then coated with gold and polished. Arrays of electromechanical devices called microshutters would allow NIRSpec to take spectra from up to 100 objects simultaneously, even if some of those objects were faint and lay next to brighter stars. Each individually controllable microshutter would be the width of a few human hairs, and NIRSpec would require more than 62,000 of them.

In addition, every piece of technology in the spacecraft would have to be engineered to endure the violent vibrations of launch, the hard vacuum of outer space and the slow cool-down to cryogenic temperatures. The telescope's optical surfaces, in particular, would have to survive all this while staying aligned to a precision of nanometres. And everything would have to perform nearly flawlessly for a minimum of five years, the baseline mission length.

Small wonder, then, that NASA ended up spending almost $2 billion just on the JWST's initial technology development. Nonetheless, the agency did not substantially cut any of the telescope's capabilities to bring

COST CURVE

The James Webb Space Telescope has consumed an ever-increasing fraction of NASA's astrophysics budget.

Astrophysics budget
JWST budget
US$ millions
Construction start
Detailed design start
Preliminary design start
1,600 / 1,400 / 1,200 / 1,000 / 800 / 600 / 400 / 200 / 0
1997 1999 2001 2003 2005 2007 2009 2011

THE JAMES WEBB SPACE TELESCOPE

The JWST, NASA's successor to the Hubble Space Telescope, will capture infrared light from the first galaxies. Too large to fit into a rocket fairing, it will unfold in orbit and cool to cryogenic temperatures.
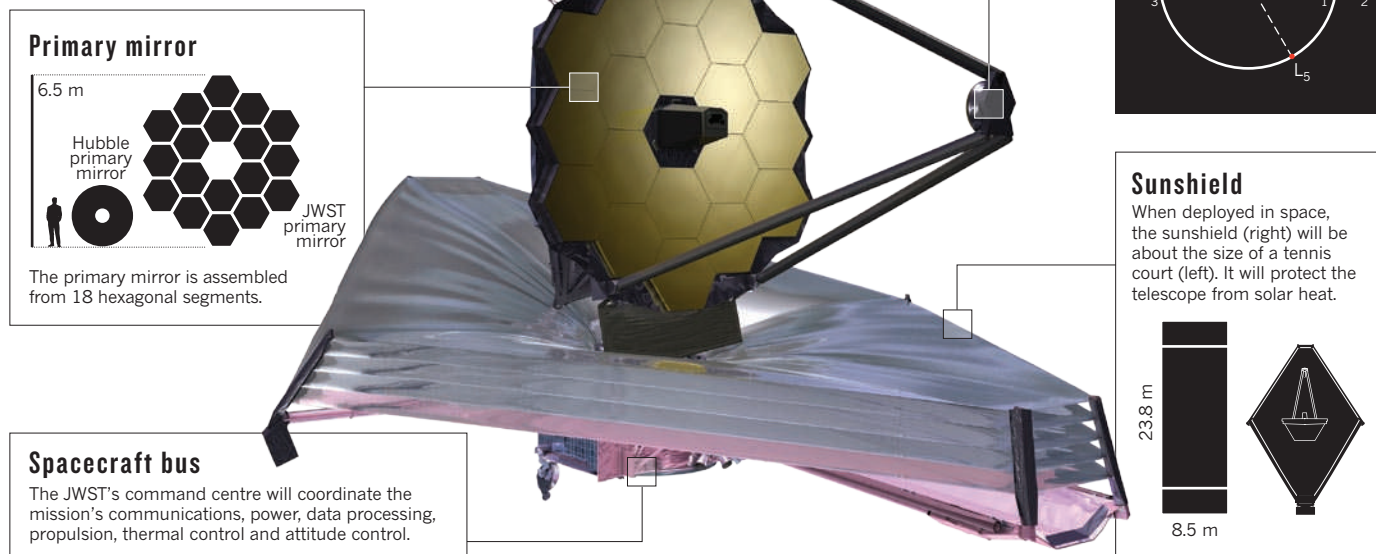
Backplane
Once the mirror has unfolded, the JWST's 'spine' will hold it still and support the telescope's cameras and spectrographs.

Secondary mirror
Light will bounce off the primary mirror into the smaller one, then to the instruments.

LAGRANGIAN POINTS
There are five places where the balance of gravitational forces allows a spacecraft to be stationary relative to the Sun and Earth. The JWST will operate opposite the Sun at the point designated L_2.

NASA

Primary mirror

6.5 m

Hubble primary mirror

JWST primary mirror

The primary mirror is assembled from 18 hexagonal segments.
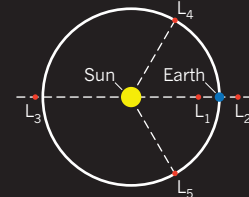
Sunshield
When deployed in space, the sunshield (right) will be about the size of a tennis court (left). It will protect the telescope from solar heat.

23.8 m

8.5 m

Spacecraft bus
The JWST's command centre will coordinate the mission's communications, power, data processing, propulsion, thermal control and attitude control.
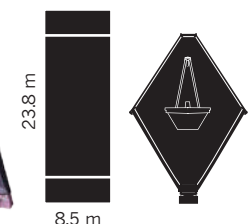
the costs back under control. Instead, it looked for partnerships, securing major contributions from the European and Canadian space agencies. NASA also maximized support for the project on Capitol Hill by awarding contracts for spacecraft components to a small army of companies and universities scattered through many congressional districts. Aerospace giant Northrop Grumman of Los Angeles, California, became the JWST's prime contractor, under NASA's Goddard Space Flight Center in Greenbelt, Maryland, which would manage the overall project.

By the time the JWST passed its preliminary design reviews in spring 2008 and NASA had officially committed to building it, the project had been transformed from its comparatively modest 'faster, better, cheaper' origins into an audacious multibillion-dollar, multi-instrument mission spanning institutions, countries and continents.

PASSING THE TEST

For nearly a year now, engineering models of the JWST's various components have been trickling into the clean room in Goddard's Building 29 for testing. (The centre's white-suited technicians can be seen at work on Internet 'Webb-cams'.) Pieces of actual flight hardware are supposed to start arriving in the same room in spring and summer 2011. All of the JWST's riskiest technologies have met their critical milestones and are on schedule for the 2014 launch.

The most substantial challenge remaining before launch is to integrate and test the flight components to ensure that they function as a whole — and, of course, to do all that without exceeding the remaining budget. NASA's traditional method is to 'test as you fly' — to operate the integrated flight hardware in conditions as close as possible to those it will experience in space. The problem is that the fully assembled telescope will be far too large to fit into any available thermal vacuum chamber. Just as the JWST's scientific objectives required new technology, mission planners have had to devise entirely new protocols to test it.

"With the JWST we have to do incremental modelling, building and testing, validating our model at each stage and then moving up to the

⟳ NATURE.COM
To learn more about the future of astronomy, visit:
go.nature.com/79ogcj

next level of assembly," says Phil Sabelhaus, the JWST project manager at Goddard. "We aren't only testing — we're also proving our ability to model correctly, which is how we will evaluate the JWST's absolute performance on-orbit." This hierarchical assembly, testing and modelling is laborious and time-consuming, more like building several telescopes than one, and is a major contributor to the JWST's remaining costs. So, unsurprisingly, it is one of the most probable targets for cost-cutting.

"There are tests that are really essential to do, and tests that would be nice to do," says Dressler. "With something of this magnitude, there is a natural tendency to double-check and triple-check, and maybe we can't afford that." On the other hand, he says, maybe they can't afford not to: it was a decision to save money on testing that allowed a defect in Hubble's primary mirror to go undetected until it was in orbit, nearly dooming the entire mission.

The JWST's supporters contend that, even with further budget overruns, the telescope will still break the historical cost pattern for large space telescopes. "Not even including its four space-shuttle servicing missions, Hubble cost $4 billion or $5 billion in today's dollars just to build and launch," Dressler notes. "Here we are, building a telescope that is almost seven times bigger, it is cryogenic, it is operating 1.5 million kilometres away, and it is costing the same amount as Hubble did, if not less. That is remarkable, and this is probably the biggest scale on which we will consider building such things in this country."

Even so, ambivalence still surrounds the JWST. Failure is not an option, either for NASA or for the astronomers it supports. Yet, in the face of flat or declining budgets, a dwindling docket of near-term astrophysics missions and rising public outrage over perceptions of runaway government spending, tough questions are inevitable. At a mid-September meeting of the agency's astrophysics subcommittee, efforts to nail down just how many extra dollars lie between the JWST and its eventual arrival at L_2 were met with silence. Until the announcement of a new budget and schedule, informed by recent panel reviews, that is the best answer anyone is likely to get. ∎

Lee Billings *is a freelance writer based in New York.*

COMMENT

M STOJKOVIC/SPL
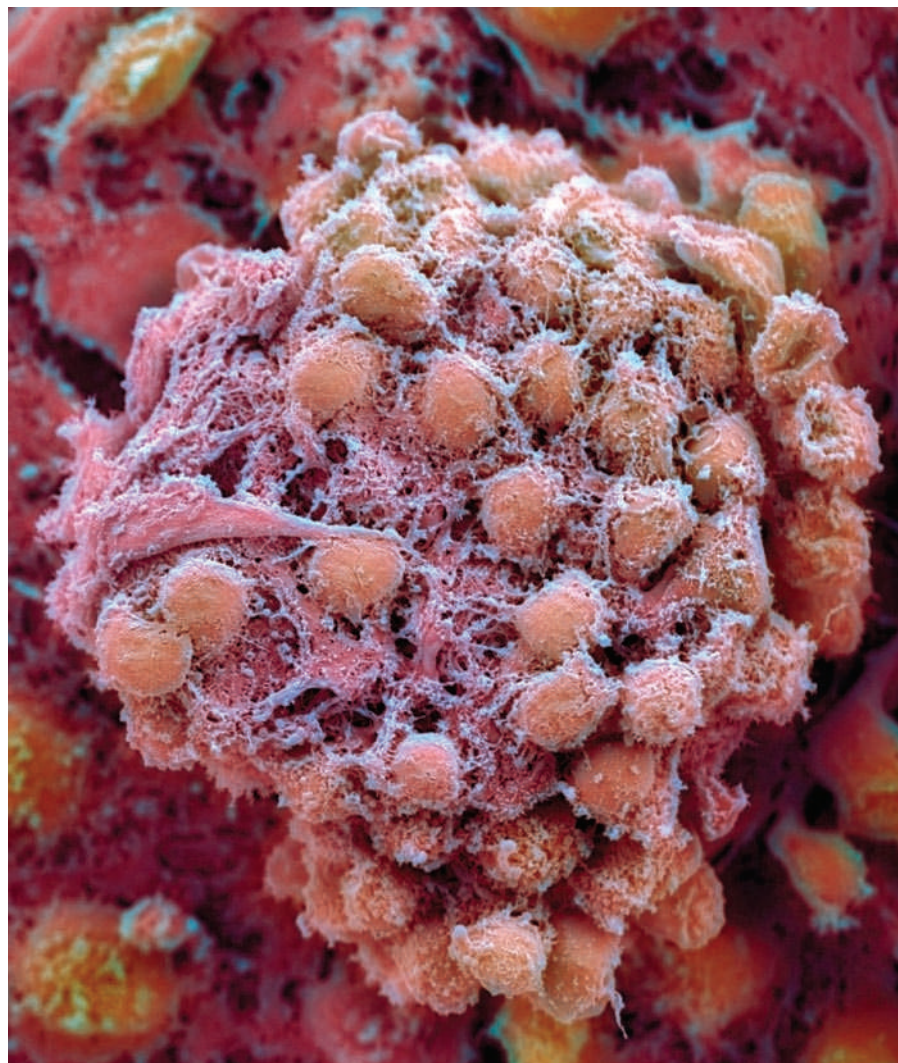
Long shadow of the stem–cell ruling

Two months on from the court decision that briefly suspended US federal funding for human embryonic stem–cell research, uncertainty still stalks the field. Here an ethicist, a team of bankers and a lawyer warn of effects of this saga that could be felt for years to come.

THE ETHICIST
Vanguard of the new biopolitics

Jonathan D. Moreno is at the University of Pennsylvania in Philadelphia.

Whatever the outcome of the legal process that has called into question the future of US federally supported human embryonic stem-cell (hESC) research, there will be no turning back the clock to the day before such funding was temporarily banned by a district court judge. Quite rightly, life scientists are wondering whether this incident signals an extended series of controversies in the United States about experimental biology.

There is a narrative that suggests that it does. Seen in the light of other incidents, and cultural and political factors, the torturous tale of hESC research in the United States is but a more emphatic example of an emerging 'biopolitics'.

The first examples of the modern politics of biology, the recombinant-DNA debate and the first human birth by *in vitro* fertilization, took place during the 1970s in a less politically fevered environment than today. Memories of the public concerns and confusion in response to those events have faded. Like stem cells, both were direct technical challenges to what many regarded as the order of biological nature, and both reminded us, as stem cells do, that the human body, for all the advantages it gives us over other creatures, shares its fundamental systems of growth, organization and reproduction with other living things. Even while airy talk of postmodernism filled the philosophy seminar rooms, over in the science buildings it was hard to deny that something pretty basic was being learned as biologists began to manipulate the underlying mechanisms of life.

There was plenty of fodder for society's doubt about the implications of science and its concerns about the hubris of scientists. These are themes that reach back to the origins of the Enlightenment, from Francis Bacon's scientist-governed utopian *New Atlantis*, to Mary Shelley's *Frankenstein*, ▶

H. G. Wells's *Island of Dr Moreau* and Aldous Huxley's dystopian *Brave New World* — all works in which the monster is not the creature, but the scientist.

But it is this stem-cell saga that has provided the fullest expression yet of the new politics of biology. Never before has a debate about a specialized laboratory practice been the occasion for passionate cultural division that surfaced in three presidential campaigns and many state elections, before completing its latest adventure in the judicial system.

Other biopolitical issues haven't achieved the status of stem cells but are based on the same competition for control. For example, a 2009 Louisiana law prohibits attempts to create, transfer or transport human–animal hybrids, and a similar bill is under consideration in Arizona; violators face prison and a seven-figure fine. Both bills were inspired by a congressional bill — drafted by the probable next governor of Kansas, Senator Sam Brownback — that seems to prohibit the use of cow eggs for somatic-cell nuclear transfer. The worry expressed by supporters of the law, that the mixing of human and animal cells tends to blur species lines and undermine human exceptionalism, is one that applies to much modern experimental biology. Britain had its own dust-up over 'cybrids' that played out in its parliament a couple of years ago.

The flashpoints of the US post-Enlightenment ambivalence about science — the abortion debate, end-of-life care, 'designer babies' and now stem cells are somewhat different from those of modern Western Europe. In the United States, genetically modified organisms are *persona non grata* on the menu. Yet the nation is the only country that was founded by a group of scientists under the explicit inspiration of the eighteenth century's valorization of reason and demonstration in the growth of knowledge. Their vision of a new nation that would be a magnet for inventors and invention was and remains embodied in the patent statute.

For much of the country's first century, anti-federalists disputed the constitutional reach of the central government in paying for 'internal improvements', including roads and bridges and innovations such as telegraphy. Although we can hardly imagine what US science and technology would look like in the twenty-first century without a robust federal role, it is remarkable that stem-cell funding is in essence tied up in a federal–state tension over internal improvements.

The United States faces a 20–30-year process of economic reconstruction that must include bio-based industries. Historically, Americans have reconciled themselves to change, however reluctantly and spasmodically, if it signified a brighter future. Without exaggerating the significance of a single policy decision, the nature of this choice foreshadows many more. Welcome to the new biopolitics.

THE BANKERS
US firms could be left behind

John M. Nolan, Emad U. Samad, Suy Anne R. Martins *and* Stephen G. Brozak *are at WBB Securities in Clark, New Jersey.*

The recent litigation in the District of Columbia Circuit attempting to suspend the public funding of hESC research in the United States also threatens privately funded research. It has created an atmosphere of grave uncertainty among Wall Street investors who now shy away from hESC products, alarmed by the increased risk that stems from protean federal policy and the ambiguous regulatory requirements (see graph).

The United States is at a crossroads. Never before has there been such a paucity of funding for the commercialization of a technology with such immense therapeutic potential. To date, we estimate that less than US$250 million has been directly committed to meaningful commercial enterprises engaged in translating hESC research into viable therapeutic candidates for human disease.

Without the immediate adoption of a clear federal policy, backed by substantial funding for basic research and product development, we believe that the market for hESC technologies in the United States will be irreparably harmed. The country will lose its position as a leading developer of regenerative medical therapeutics despite the fact that as many as 60% of Americans now approve of the creation of hESC lines

for research and therapeutic use[1].

Researchers and companies are already turning to other nations to advance basic hESC science and product development[2]. The United Kingdom, for example, has made hESC research a national priority, with funding commitments in excess of £350 million ($556 million) and economic incentives that have already lured many top researchers to the country. Government-sponsored programmes, such as the UK Stem Cell Initiative, have encouraged collaborations between public and private institutions, in some instances mandating academia to seek out partners in industry for projects to qualify for government funding[3].

By comparison, only $42 million of the National Institutes of Health's (NIH's) roughly $30-billion budget in the 2007 financial year was allocated to hESC research. Even after President Barack Obama lifted the Bush-era cell-line restrictions, federal funding levels increased to a projected $123 million in 2010, far less than the allocations for many areas such as nutritional education, alcoholism, substance abuse and gene therapy. Compared to the $424.8 million allocated to the Human Genome Project in 2000 ($335.9 million by the NIH and $88.9 million by the Department of Energy) and the roughly $2.6 billion that was allocated to the project throughout the 1990s, current funding levels for hESC research are simply not sufficient to bring a concept from inception to commercialization, nor have they been adequate to entice private industry into the market.
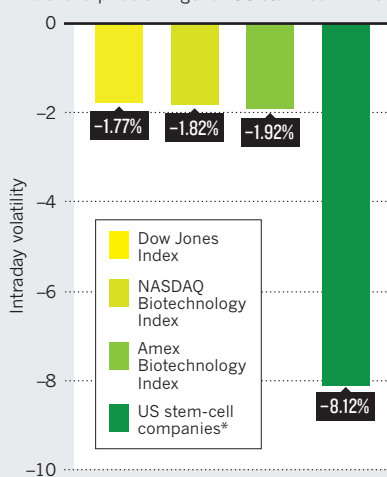
The United States must act now to rectify the missed opportunities of the past decade and to protect its future scientific, medical and commercial interests. It can begin by revising the 1996 Dickey–Wicker Amendment to permit future and continued use of embryonic cell lines.

We also recommend that the US government makes a financial commitment as large as that dedicated to the Human Genome Project and increase yearly NIH appropriations for hESC research to at least $500 million. Otherwise, as research continues elsewhere, European pharmaceutical companies will continue to build a strong intellectual-property position that they will use to protect their investments and generate perpetual development and revenue cycles.

Some US companies have built substantial hESC intellectual-property portfolios. However, their science and commercialization pipelines are not maturing at the same pace as those of their European or Asian counterparts. Thanks to scant national coherence and significant regulatory risk, the US capital markets have failed to provide financing in sufficient sums to spur serious product development. As a result, hESC science and technology is now concentrated in the hands of a few undervalued US companies.

INDISCRIMINATE EFFECT

On 23 August, the suspension of funding for human embryonic stem-cell research caused wild share-price swings for US stem-cell firms.

- Dow Jones Index: −1.77%
- NASDAQ Biotechnology Index: −1.82%
- Amex Biotechnology Index: −1.92%
- US stem-cell companies*: −8.12%

Intraday volatility

*Average of ten adult- and embryonic-stem-cell companies

SOURCE: BLOOMBERG

Over the past two years, growing numbers of pharmaceutical companies from emerging economies have vied for entry into Western pharmaceutical markets by manufacturing generic drugs. China, for example, is poised to become the world's third-largest pharmaceutical market next year and will contribute the same in annual sales in 2013 — more than $40 billion — as the US market. Meanwhile, American and European pharmaceutical companies have become desperate to sustain eroding revenue as proprietary patents for blockbuster drugs expire, allowing more generic competition.

To corner the market that may hold the next medical revolution, an Asian pharmaceutical company could easily decide to acquire US companies that have advanced technologies but very low market valuations. If foreign pharmaceutical companies focused resources, they could proceed with product development at a pace that the US pharmaceutical industry would be unable to match. Such a move would signify a shift in the balance of power of the health-care market and set US stem-cell science back a generation.

1. Gallup stem cell research poll; available at go.nature.com/y5kxvi
2. Sipp, D. *Regen. Med.* **4,** 911–918 (2009).
3. UK Stem Cell Initiative (UKSCI) *UK Stem Cell Initiative: Report and Recommendations* (2005).

THE LAWYER
Why US science is stuck in the dock

Patrick L. Taylor *is at the Petrie–Flom Center for Health Law Policy, Biotechnology, and Bioethics at Harvard Law School in Cambridge, Massachusetts.*

The judge forgot the potential for cures, writes one editorial. Appeal the decision, pass a new statute! But the impact of the court's methods will linger long after the dust has settled. The implication that no facts are certain in the United States means that no science is safe.

The court had to interpret the Dickey–Wicker Amendment, a budget rider disallowing funding of research in which human embryos are "destroyed, discarded, or knowingly subjected to risk of injury or death greater than that allowed for research on fetuses in utero". Sound court orders depend on sound determination of two kinds of facts. The first is objective: will it cause harm to stop funding immediately? (No, said the court, without consulting other extramural researchers.) Whose harm will be greater? (Continued funding would seriously harm two plaintiff researchers claiming potential competitive

injury to their non-hESC research, said the court, whereas stopping all hESC funding will cause no harm, and preserve the status quo, because hESC researchers can go to industry.) The court said a stop-order was consistent with the "public interest", but didn't say why — despite overwhelming public support for hESC funding.

The second kind of fact is interpretive: what did Congress mean, and what did it want? The 'Chevron' ruling, named after the Supreme Court case announcing it, requires courts to stick to legal text if it's unambiguous, as that best fulfils congressional intent. If a law is ambiguous, courts must defer to agencies charged by Congress to administer it.

US law is filled with useful heuristic rulings, establishing methods or reconciling new developments with old categories. But if misapplied or too crude, these rulings can supplant justice, prevailing over what basic factual inquiry would have required. Before slavery was abolished in the United States, courts were asked whether African people were property rather than persons. Yes, said the courts, so laws of sales and inheritance swung into place, paving the path from slavery to slums with falsehoods.

The district court's decision was an ingeniously literal use of Chevron. It capitalized upon the requirement to stick to law alone if the law is clear by determining that Dickey–Wicker is "unambiguous". So the court could exclude evidence of congressional and presidential activity conclusively mandating hESC research funding, and could decide that all research using hESCs is of a piece. The differences between research that derives and research that uses hESC lines are well established. Congress is aware of them. Regulations, agency guidance and science practice would have shown that research protocols rarely encompass the creation of ingredients — cells, drugs and reagents are provided by third parties. A study that involves injecting hESCs to cure neonatal paralysis will raise important ethical and scientific questions. But it will not be research in which a human embryo is "destroyed".

Such a broad reading of what it means for research to involve destroying embryos threatens important research. By the same logic, could federally funded research on HeLa cells now be construed as 'research killing a patient', because Henrietta Lacks died from the cancer that was the source of the original cells? Could research to correct fatal heart syndromes in fetuses, or all research into genetic diagnostic tests also be imperilled? More crucially, a judicial finding of "unambiguity" — which facts would have rebutted — now permits courts to ignore the NIH and other agencies, and scientists who engage with

Congress to influence legislation.

In a way, this was a legal accident waiting to happen. From the 1990s, political debate about stem cells has been excessively influenced by Dickey–Wicker's emphasis on what government would fund. Ethical rules linked to NIH funding — addressing issues such as the sharing of data or materials — did not apply to most stem-cell research because it was not federally funded. The result was complex funding rules, fear in the research community and patent monopolies.

Yet in this ethics vacuum, something spectacular occurred: people thought about the questions publicly, debated them closely and reached a reasonable, nuanced conclusion. They saw what other countries, such as the United Kingdom, did. The media established an ongoing conversation across international borders. Scientists and others created, through national and global guidelines, a self-regulatory ethical framework that did what laws did not — such as requiring independent review to evaluate scientists' proposals, barring research on embryos once nervous-system development has begun, prohibiting coercion of egg donation and forbidding financial inducements for research eggs and embryos. Global discussion led to a shared US vision of ethically permissible funding. Subsequently, the NIH introduced rules that accurately reflected popular will and an interpretation of Dickey–Wicker that Congress had repeatedly confirmed.

"For public ethics to become public law, we need to know when law fails, and why."

The suspension saga has effectively annulled the marriage of law and ethics embodied by the final NIH rules. Public ethical consensus, votes conscientiously considered and norms for open science became irrelevant. Legal fictions replaced facts, and a heuristic legal ruling designed to respect congressional and public will was the very instrument of democracy's defeat.

Now the branches of government must work together not just to fix hESC funding but to stamp out the methods used to bring it so low — to head off future damage to novel science. Judicial appointments also need examining. They should not be principally based on divining candidates' personal politics, but more on the choice to set personal politics aside. How candidates discern fact, understand Congress and reconcile law with what is new, are key. Congress must also close the loopholes allowing courts to ignore authoritative evidence of congressional intent and textual ambiguity.

We need a new watchdog that tells us when law radically misaddresses science's rapid developments. For public ethics to become public law, we need to know when law fails, and why. And then we must act. ∎

↻ NATURE.COM
See the stem-cell injunction special at
go.nature.com/ljonzx

AUTUMN BOOKS

ILLUSTRATIONS BY G. POTENZA

**COSMOLOGY**

Space–time turn around

Lee Smolin marvels at Roger Penrose's masterly and imaginative argument that our Universe is one of a succession.

No living physicist has yet made a discovery as great as those of Isaac Newton or Albert Einstein, but Roger Penrose is in a better position to do so than most. Combining a mastery of mathematics with trust in his own research compass, Penrose — a mathematical physicist at the University of Oxford, UK — is driven by a heroic obsession to understand fundamental puzzles about nature. The depth of his thinking and fertility of his creativity concerning the mathematical foundations of modern physics place him above his peers.

In *Cycles of Time*, Penrose introduces his most outrageous and subtle idea yet. Answering the question of why the future is so different from the past — why eggs crack into pieces that never spontaneously reassemble, for example — he lays out his thinking on the origin and fate of the Universe. Penrose addressed this problem in his first popular-science book, *The Emperor's New Mind* (Oxford University Press, 1989). His latest volume describes a new way of resolving that problem. It is an astounding idea, which, if true, would revolutionize physics and cosmology.

We should pay attention because Penrose has repeatedly been far ahead of his time. The most influential person to develop the general theory of relativity since Einstein, Penrose established the generalized behaviour of space-time geometry, pushing that theory beyond special cases. Our current understanding of black holes, singularities and gravitational radiation is built with his tools.
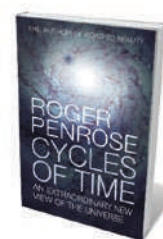
**Cycles of Time:
An Extraordinary
New View of the
Universe**
ROGER PENROSE
*Bodley Head: 2010.
320 pp. £25*

His work in the 1960s on quantum gravity has borne dramatic fruit within the past five years. Penrose introduced two influential concepts: spin networks, which in 1988 seeded an approach called loop quantum gravity; and twistor theory, a recasting of space-time geometry that has generated a recent breakthrough in our understanding of gauge theories, the basic ingredients of the standard model of particle physics.

Readers will not be disappointed with the audacious ideas in his latest book. It starts with a masterful explanation of the directionality of time. A gifted popularizer of science, Penrose skilfully breaks the normal rules by including equations and describing subtleties and uncertainties. He is honest too, clearly distinguishing established science from his own speculations, and relating opposing views and alternative ideas with balance.

Penrose then sets out his proposal. It rests on the puzzle that the apparent initial state of the Universe is highly improbable — a quandary he has emphasized for years. By running the laws of physics backwards from the Universe's present state, we can work out what it looked like just after its birth. But given all of the possibilities conjured up by physics, it is extremely unlikely that a randomly picked universe will resemble our own.

The initial state of our Universe is special, Penrose argues, because it is simultaneously very hot and very cold. The matter and electromagnetic radiation are exceedingly hot, at a temperature that approaches infinity as we go back in time to the singularity of the Big Bang. But because there is no energy in gravitational waves, he says, the geometry of space-time has a temperature of essentially zero. Both extremes mean that we can simplify our description of the state of the Universe.

COOL GEOMETRY
At extremely high temperatures, the elementary particles that comprise matter and radiation are indistinguishable and their interactions negligible because their energies are tiny compared with the Universe's heat. The newborn Universe is essentially a hot gas of photons, and everything that happens to that gas is determined by one number: its temperature. The coldness of the space-time geometry also means that we can simplify its structure — at zero temperature there are no black holes and space is uniform.

Penrose argues that the direction of time is explained by the evolution of the Universe from this special, simple and improbable state to more probable ones. The unfolding of increasing numbers of random events drives the arrow of time. This is an expression of the familiar second law of thermodynamics that randomness — or entropy — tends to increase. The problem of explaining the arrow of time is then reduced to the question of why the early Universe was so special.

Penrose tries to answer this by turning from the very early Universe to its extreme future. As it expands, the density of

matter — and hence energy from ordinary stuff — wanes. But the 'dark energy' associated with the vacuum of space remains constant (at least in simple models of it) and eventually dominates. Dark energy accelerates the expansion, further diluting the matter. All black holes will evaporate and any other space-time features will be ironed flat by the exponential expansion. Stars and galaxies will dissemble if, as Penrose postulates, elementary particles eventually decay to photons and other massless particles.

If these hypotheses are true, then at very late times the Universe will look a lot like it did at very early times — its spatial geometry is homogeneous and flat, and it is filled with a gas of photons. There is one difference: the temperature and density of the early

> IT IS POSSIBLE THAT OUR EARLY UNIVERSE IS THE LATE UNIVERSE OF A PREVIOUS ERA. THIS IS PENROSE'S BIG IDEA: **DELICIOUSLY ABSURD,** BUT JUST POSSIBLY TRUE.

Universe differ by an enormous factor from its end point. This can be understood as a change of scale, such that an act of compression — by a vast factor — could turn the late Universe into the early one.

Penrose pulls one more trick out of his hat: the insight that physics in both the early and late regimes is insensitive to scale. Briefly, this is because massless particles move at the speed of light, at which point time stands still for them. Because there is no clock ticking, there is no reference against which they can measure a scale of length or time.

So if the only difference between the very early and late Universe is scale, and physics in both of these extremes is insensitive to changes of scale, then it is possible that our early Universe is the late Universe of a previous era. This is Penrose's big idea: deliciously absurd, but just possibly true. Moreover, it doesn't matter if such a transition took an eternity — photons are insensitive to the passage of time.

Penrose's concept joins several other proposals, such as loop quantum cosmology, that replace the Big Bang singularity and allow time to run before the Big Bang occurred, suggesting our Universe is the progeny of a previous one. Other ingenious mechanisms for making the history of the Universe cyclic — so that it repeatedly swells and contracts — have been proposed by physicists Paul Steinhardt and

Neil Turok and their colleagues. But these exotic proposals involve theories of quantum gravity, which Penrose has no need for in his hypothesis.

INFLATION POPPED
Penrose's proposal has another advantage, in common with other hypotheses that eliminate the singularity. It suggests that before the Big Bang, there would have been plenty of time to set up the correlations seen in observations of the cosmic microwave background and distributions of galaxies. Consequently, there is no need for the hypothesis of rapid inflation of the Universe very early in its history. This is potentially a good thing, because inflation is hard to stop once it is started, and can easily lead to a multiverse with an infinite number of universes like our own.

The multiverse scenario raises challenges because the explanation for why our Universe is like it is must then rely on untestable assumptions about an infinite ensemble of unobservable universes. This in turn raises puzzles about applications of probability, and requires use of the anthropic principle — further decreasing the empirical content of the theory. The anthropic principle posits that our Universe is one among a vast ensemble, most of which cannot contain life. Because one is free to make arbitrary hypotheses about the other universes, which are neither observable nor need be like our own, almost any property of our Universe can be explained away. All of these problems are avoided by hypotheses such as Penrose's that invoke a succession of universes rather than an unobservable infinite simultaneous plurality.

Despite this, inflation has so far proved successful in accounting for the observed patterns in the cosmic microwave background. The challenge of scenarios of succession such as Penrose's is to account for those observations and make a prediction that differentiates it from inflation. Then experiment can decide. Penrose's proposal therefore needs development and reflection as a scientific idea.

Cycles of Time starts off as a masterpiece of pedagogy and becomes more challenging as the book progresses. But it is worth reading to see Penrose's extraordinary mind working to confront one of the fundamental puzzles of our present understanding of the Universe. ■

Lee Smolin *is a faculty member at the Perimeter Institute for Theoretical Physics, Waterloo, Ontario N2L 2Y5, Canada, and author of* The Trouble with Physics.
e-mail: lsmolin@perimeterinstitute.ca

NEUROSCIENCE

Learning to see

Steve Silberman is moved by Oliver Sacks's poignant account of losing his vision through cancer.

Eight days before Christmas Day 2005, neurologist Oliver Sacks — author of *Awakenings* (1973), *The Man Who Mistook His Wife for a Hat* (1985) and other popular collections of case histories — went to the movies. Sitting in the dark theatre, he noticed an incandescent flickering to his left, which he took to be the first signs of a migraine. But as a blind spot appeared and grew, the 77-year-old physician started to panic.

When the floor lights pointing to the exit abruptly vanished, he burst out of the cinema and phoned a colleague, who urged him to see an ophthalmologist. The diagnosis was sobering: Sacks had a melanoma in his left eye that would require prompt treatment. Thus the neurologist took his first steps on a harrowing course of transformation, mirroring those of his patients. *The Mind's Eye* is Sacks's frank and moving account of that journey.

Sacks has written about neurological disorders — such as autism, colour blindness
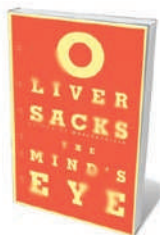
The Mind's Eye
OLIVER SACKS
Alfred A. Knopf/
Picador: 2010.
288 pp/272 pp.
$25.95/£17.99

and synaesthesia — as a way of talking about the higher orders of the human mind since he published *Migraine* in 1970. At a time when the brain's plasticity was barely acknowledged in medicine, Sacks saw its reparative power in the lives of his patients, guiding them toward wholeness and vitality after a traumatic loss of ability. Defects, disorders and diseases, Sacks wrote in *An Anthropologist on Mars* (1995), can have a paradoxical role, "by bringing out latent powers, developments, evolutions, forms of life, that might never be seen, or even be imaginable, in their absence".

In *The Mind's Eye*, Sacks probes visual dysfunctions — including alexia (an inability to make sense of words), prosopagnosia (a failure to recognize faces) and his own ocular melanoma — to examine the complex roles of sight in human life and the constitution of personal identity.

He considers the case of Lilian Kallir, a concert pianist who became increasingly unable to make sense of her world visually. She lost the ability to read musical scores because of a rare degenerative condition called posterior cortical atrophy. Many elements of Kallir's story will be familiar to readers of Sacks's other books: her letter to the doctor seeking advice of last resort, the elusive diagnosis and the lofty cultural milieu of the patient. Also familiar are Sacks's attempts to comprehend the scope of Kallir's condition by visiting the vivacious 67-year-old musician at home.

Part of the appeal of Sacks's books is his depiction of an idealized world of thoroughly personalized medicine. Few physicians have the time or inclination to make house calls any more. Fewer still would say to a visually impaired patient, as Sacks does, "Let's go out, let's wander" — and then dress in red so that the patient can spot him in the bustling crowds of Manhattan.

This is not merely Sacks showing off. One of his role models, the late French neuropsychiatrist Jean Lhermitte, advised accompanying patients to a bistro to observe how they were coping with their illness. After Sacks visits the apartment that Kallir shares with her devoted husband, he writes about the ad hoc methods that the couple devised to make the pianist's illness less disabling. In the kitchen, for example: "Things were categorized not by meaning but by color, by size and shape, by position, by context, by association, somewhat as an illiterate person might arrange the books in a library. Everything had its place, and she had memorized this."

Like most of Sacks's case studies, Kallir's story does not come to any satisfying therapeutic resolution. There is no breakthrough, no wonder drug and no hope of lasting recovery. But the ability of the pianist and her husband to maintain a shared sense of continuity in increasingly disordered circumstances is testimony to the resilience that is Sacks's

---

NEW IN PAPERBACK

Highlights from this season's latest releases

The Art and Politics of Science
Harold Varmus (W. W. Norton, 2010; $15.95)
In his memoir, Nobel prize-winner Harold Varmus reflects on his work in cancer biology, his directorship of the US National Institutes of Health and the many political battles that he has fought over science. His ability to connect basic research and medical application is evident. "Varmus reveals a sharp, analytical intelligence as well as great enthusiasm for his work and profession", wrote reviewer Iain Mattaj (*Nature* **458,** 32; 2009).

overarching theme. Rather than being about disease, his tales are more about his patients' astonishing capacities to adapt — and even thrive — in radically transformed worlds. His books resonate because they reveal as much about the force of character as they do about neurology.

The Mind's Eye also relates how an Australian psychologist named Zoltan Torey, rendered blind at 21 by a splash of acid, cultivates his photographic memory to the point that he shocks his neighbours by replacing the gutters of his house alone at night. In another chapter, Canadian novelist Howard Engel discovers that his morning *Globe and Mail* has been rendered into Cyrillic or Korean; it is his brain, of course, that has been translated by a stroke. After years of exhausting effort to engage language in new ways — composing by dictation, learning to scan words by linking adjacent letters — the novelist teaches himself to write books again.

For Sacks, disorders of vision, including his own, open a window on the brain's surprisingly active role in the authoring of experience. While under treatment for the ocular melanoma, the neurologist undertook a series of fascinating self-experiments. In one exercise, for example, he tested the limits of his brain's ability to fill in temporary gaps in his visual field caused by radiation treatment. Sacks found that repetitive patterns such as brickwork, and even clouds and trees, readily appeared to preserve the illusion of a seamless panorama around him. Faces, however, were beyond the conjuring ability of his visual cortex. "I've learned that the brain is always busy," he told me in an interview last summer.

Thankfully, Sacks's tumour has not returned, but he is still learning to cope with the aftermath, including a possibly permanent loss of three-dimensional vision — a poignant turn of events for a proud member of the New York Stereoscopic Society.

To maintain his own sense of continuity in the face of these challenges, Sacks will have to draw inspiration from the patients he has written about for 40 years. "The problems never went away," he quotes Engel as saying, "but I became cleverer at solving them". ∎

Steve Silberman *is a writer based in San Francisco, California.*
e-mail: digaman@sonic.net

Hitchers, outcasts and wasteland beauties

Sandra Knapp revels in a portrait of weeds as resilient rebels shaped by our meddling with the wild.

Like humans, weeds are pervasive, domineering and badly behaved. But they adopt these traits only in order to reproduce. As naturalist Richard Mabey explains in *Weeds,* they are an in-your-face example of evolution by natural selection: weeding benefits weeds by allowing those that evade the hoe to produce seeds that inherit the very characteristics that allowed escape; using herbicide causes weeds to become more resistant to such poisons.

Mabey weaves social history, psychology, literature and art into his clear rendering of plant biology. Explanations of evolution sit alongside explorations of flower symbolism in Shakespeare. This blend, familiar to fans of his earlier reflections on nature in the wild, broadens the book's scope to human attitudes to plants in general.

Indeed, the concept of a weed makes sense only in relation to people — they are plants that cause us trouble by growing where we don't want them. Most of the social connotations of weeds are negative: unruly, weak or aggressive. Yet these designations are fluid. Some plants, such as St John's wort (*Hypericum perforatum*) or hemp (*Cannabis sativa*), have passed from love to hate and back again. Others, such as autumn ladies' tresses (*Spiranthes spiralis*), are a rampant but admired invader of our lawns.

Some weeds considered ubiquitous today were once rare: rosebay willowherb (*Epilobium angustifolium*), depicted among the fine flora on the ceiling of the Natural History Museum in London, was described by some nineteenth-century botanists as a woodland plant 'not often met with in the wild state'. This magenta-flowered perennial carpeted the bombed areas of 1940s London, earning it the common name of fireweed. Its

Weeds: How Vagabond Plants Gatecrashed Civilisation and Changed the Way We Think About Nature
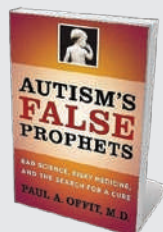RICHARD MABEY
Profile Books: 2010. 288 pp. £15.99

tiny seeds, carried on downy plumes, were dispersed by turbulence along railways; it now colonizes cities across Europe and North America. It is a good example of how weeds are a human construct, promoted by our tendency to disturb land.

Naturally invasive or easily transported species are also troublesome, particularly on islands with rare flora such as Hawaii, the Galapagos and Australia. For example, the velvet tree (*Miconia calvescens*) has taken over rainforest areas in Tahiti and is spreading on Hawaii; it chokes off native vegetation, preventing natural forest regeneration in these fragile habitats. But these plants arrived with people. *Homo sapiens* is the ultimate invasive species — coming out of Africa to colonize the globe, altering the planet beyond recognition.
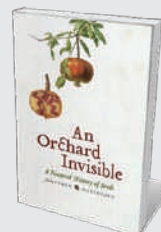
Weeds highlights our ambivalence about naturalness and artificiality. We often think of pristine nature as the landscape we, or our grandparents, grew up with. Yet nature changes all the time. In the Pleistocene, much of northern Europe was covered with ice: no plants grew. Our entire flora is invasive, but that hasn't stopped us loving it. ∎

Sandra Knapp *is a botanist at The Natural History Museum, London SW7 5BD, UK.*
e-mail: s.knapp@nhm.ac.uk

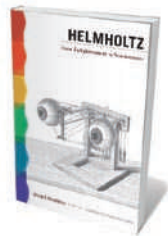HISTORY

A polymath rediscovered

George Rousseau uncovers the physiological side of Hermann von Helmholtz.

When the Harvard University psychologist Edwin Boring dedicated his classic 1942 monograph *Sensation and Perception in the History of Experimental Psychology* to Hermann von Helmholtz, many American readers wondered why. Helmholtz was a German, the Allies were rallying against the Nazi menace, and the United States had just entered the war. Few beyond professional historians of science knew about Helmholtz's work.

Boring justified his choice: "There is no one else to whom one can owe so completely the capacity to write a book about sensation and perception." Sixty years on, Helmholtz's major contributions to physiology and medicine, including his theories of visual and aural perception, have been largely eclipsed by his work in physics. In *Helmholtz*, neuroscientist Michel Meulders redresses the balance, showing that this towering figure was as influential as philosopher Immanuel Kant and as visionary as polymath Johann Wolfgang von Goethe.

Part of the reason for Helmholtz's partial invisibility today is that he wrote in German. It took decades for his work to reach the English-speaking world; his *Popular Lectures on Scientific Subjects*, delivered in the 1850s, were translated in the 1870s and 1880s, and his acoustical masterpiece, *On the Sensations of Tone as a Physiological Basis for the Theory of Music* (1863), in 1885. After this flurry of works — distributed during Helmholtz's last two decades — came tributes on his death in 1894. His Jewish student Leo Koenigsberger published a classic biography, again in German, in 1902, which was translated into English in 1906.

An extensive obituary in the 1896 *Proceedings of the Royal Society of London* portrayed Helmholtz as the most important physicist of the epoch. His work on the conservation of

Helmholtz: From Enlightenment to Neuroscience
MICHEL MEULDERS (TRANSLATED BY LAURENCE GAREY)
The MIT Press: 2010. 264 pp. $27.95, £20.95

energy that led him to formulate the first law of thermodynamics in 1847 was widely cited — electromagnetism was cutting-edge science. But interest in his physiology and medicine was lost. Helmholtz himself pursued physics more than physiology after the 1870s, and his theories of sight and sound were bitterly contested well into the twentieth century. Meulders restores Helmholtz's legacy by placing him within the history of science and by locating him as an aesthetic thinker as well as a scientist.

A welcome and surprising inclusion in the book is Helmholtz's role within the aesthetics of music. Meulders is right to retrieve this overlooked aspect — only a handful of specialized monographs have touched on it before. Helmholtz tackled the aesthetics of pitch and tone in 1857, after a century of neglect. "Music has hitherto withdrawn

MEULDERS DOESN'T SURRENDER HIS ADMIRATION — AT TIMES VERGING ON HERO WORSHIP.
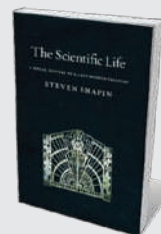
itself from scientific treatment, more than any other art," he wrote. Poetry, painting and sculpture borrow from the world of experience, he explained, but music seems to "reject all anatomization of pleasurable sensations".

Helmholtz developed a 'resonator' device, a pierced sphere of glass or brass with a narrow neck, to demonstrate musical pitch and tonal

colour. His view was that music depends on human experience and on the physiology of the senses for its effects. Helmholtz's physiological theory of music had a lasting impact on the composers Alexander Scriabin and Nikolai Rimsky-Korsakov, and on many twentieth-century academic musicologists.

Meulders brings in other German intellectuals on whose work Helmholtz built. For example, he analyses the theory of physics and physiology of colours published by Goethe as *Zur Farbenlehre* (Colour Theory) in 1810. Yet Goethe does not come to life in the book in the same way as Helmholtz's teacher Johannes Müller, portrayed as a genius who overcame insomnia and depression to hew a science of physiology.

Müller demonstrated in his famous Berlin laboratory that "the results of all physiological research must be, in the end, psychological in nature". Small wonder, then, that he assigned to his protégé Helmholtz a doctoral thesis topic in the 1830s based on invertebrates in Müller's own collection, which was eventually published as *Nerve Fibres Arising from the Ganglion Cells Discovered in 1836*. In this, Helmholtz built on the ideas of his teacher to bring together physiology and psychology.

Yet curiously, Meulders writes, Helmholtz never referred to the brain. My main reservation is that the book does not unpack this statement. Helmholtz consistently ignored anatomical data on the nervous system, and probably mistrusted the concept, popular at the time, that anatomical and psychological processes were identical. Thus he did not link the psychology of perception with the physical brain, and bought into an older theory of mind, with the soul as the arbiter of the senses. Helmholtz's defiance of nineteenth-century natural philosophy through his enduring omission of the brain is strange, and I hope another author will pursue it.

Imperial Nature: Joseph Hooker and the Practices of Victorian Science
Jim Endersby (Univ. Chicago Press, 2010; $25)
Botanist Joseph Hooker became one of the first professional scientists when research began to be funded by governments. "A refreshing record of how scientists worked during this transition," wrote Sandra Knapp (*Nature* **453**, 721; 2008).

The Scientific Life
Steven Shapin (Univ. Chicago Press, 2010; $20)
Historian Steven Shapin shatters myths about the divide between pure and commercial science by arguing that moral values are as abundant in industry as in academia. Reviewer Jerome Ravetz described it as "required reading for all scientists" (*Nature* **457**, 662–663; 2009).

Meulders concludes his book with three incisive chapters on the aesthetics of music. In one he deals with the Pythagorean legacy, especially the idea that mathematical relationships were the basis of harmony and tone. In the second he considers 'the musical ear', demonstrating that findings in auricular physiology, particularly Italian anatomist Alfonso Corti's discovery in 1851 of fibres that function as acoustical sensor cells in the cochlea, had complicated the aesthetics of sound.

This chapter is a triumph of compression of a vast province of physiology and aesthetics into a few pages. Surveying the musicological terrain from the argument between Jean-Philippe Rameau and Jean le Rond d'Alembert to Johann Sebastian Bach and Andreas Werckmeister, and on to Mozart and Mendelssohn, Meulders pauses to explain how Helmholtz the empiricist understood music theory and aesthetics as a grand unifier. Musical sounds, he thought, can only be understood as great art by combining anatomy, physiology, philosophy and psychology. The third of these chapters meditates on Helmholtz's nostalgia, intuition and memory — an odd amalgam, the breadth of which adds to Meulders's claim for Helmholtz's genius.

Meulders stitches together the thoughts of a lifetime into his slim book. He doesn't surrender his admiration — at times verging on hero worship — despite the occasional critique. The approach is hit-and-miss and does not amount to the much-desired extended interpretation unifying Helmholtz's physiology and aesthetics, but it is a brave start.

Meulders sums up his subject thus: "With his will to unify so many different scientific disciplines in a coherent entity, he proved once again his veritable gluttony for science and knowledge." Some may find Meulders equally gluttonous, but his book demonstrates that Helmholtz was indeed a polymath *par excellence*. ∎

George Rousseau *is a professor of history and co-director of the Centre for the History of Childhood, University of Oxford, Oxford, OX1 4AU, UK. He is author of* Nervous Acts.
e-mail: george.rousseau@magd.ox.ac.uk

ECOLOGY

Conservation thriller earns its stripes

A travelogue about tiger poaching in Russia's far east opens up a new genre, discovers **Geoff Marsh**.

Yuri Trush steadily points his camera at the stubs of bone protruding from a pair of thin rubber boots lying in the blood-speckled snow. As the leader of an Inspection Tiger anti-poaching unit, his job now is to piece together the details of Vladimir Markov's run-in with the tiger. Judging by the whimpering of Trush's dog, the big cat in question remains close by, among the trees.

Inspection Tiger is a government agency that was set up to combat poaching in Primorskiy Kray (or Primorye) — an area the size of Washington state in the far east of Russia, bordered by China and North Korea. Trush's team travels in a decommissioned army truck, armed with knives, pistols and semiautomatic rifles. Their mission is to intercept poachers and to resolve
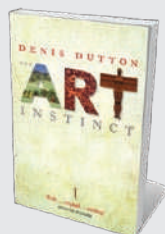

The Tiger: A True Story of Vengeance and Survival
JOHN VAILLANT
Sceptre/Alfred Knopf: 2010. 352 pp. £18.99/$26.95

the locals' conflicts with the largest cats in the world.

In *The Tiger*, author John Vaillant relates his travels across the region while investigating the pressures on tiger conservation. His vivid portrayal of Primorye reveals a unique ecosystem at the crossroads of four distinct biomes: the Siberian taiga forest, the steppes of Mongolia, the subtropics of Manchuria and the boreal forest of the far north. A peculiar mix of hardy alpine
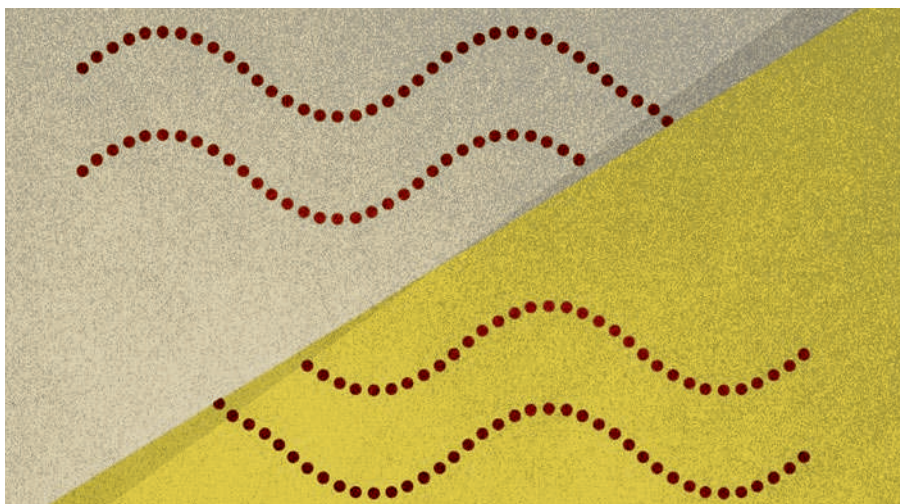

The Art Instinct
Denis Dutton (Oxford Univ. Press, 2010; £9.99)
Art appreciation has an evolutionary basis, according to philosopher Denis Dutton. The basic elements of aesthetic taste are similar across cultures and are part of our evolutionary heritage rather than being socially constructed, he claims provocatively.

Pink Brain, Blue Brain
Lise Eliot (OneWorld, 2010; £12.99)
Neuroscientist Lise Eliot marshals the latest evidence to show that social pressures are the main cause of behaviour differences between boys and girls. Although small gender variations are apparent at birth, they grow as our plastic brains quickly become modified by experiences.

QUANTUM PHYSICS

Tripping the light fantastic

Geoff Pryde on the weird world of quantum entanglement.

The only way to understand the quantum world is to measure it. This empirical view is dear to the heart of Anton Zeilinger, now at the University of Vienna, a leading figure in quantum physics through his work on correlated photons. In *Dance of the Photons,* he explores the phenomenon of quantum entanglement, the quantum correlations in the properties of particles.

When two photons are made to interact, they share their quantum information and become 'entangled'. If one travels off, it retains knowledge about its counterpart. So measuring one can determine the state of the other, even if they are far apart. Albert Einstein was worried by such reasoning: instant messaging between entangled particles contradicted his theory of relativity, which stated that signals cannot travel faster than the speed of light, unless you allow the crazy idea that particles do not have real properties independent of measurement. Quantum mechanics, he

decided, was not up to explaining the world.

Zeilinger explains that Einstein was wrong. Experiments in the 1980s and 1990s proved the weird predictions of quantum entanglement to be true. Putting the reader in the role of discoverer, he describes these tests through the eyes of fictional students Alice and Bob, namesakes of the characters regularly put to work in explaining quantum physics. Examining the philosophical and technological implications of spooky quantum phenomena, he points to big issues that demand further thought — the inherent randomness of quantum physics and the role of the observer in determining a quantum particle's reality.

As well as giving an overview of other work, Zeilinger relates in detail his own group's research. For instance, he describes a 'delayed choice entanglement swapping' experiment he has carried out using four photons (1, 2, 3, 4). Two pairs share prior information: photons 1 and 2 are entangled, photons 3 and 4

are also entangled, but there is no correlation between those pairs. Making a particular type of quantum measurement — known as a Bell measurement — jointly on photons 2 and 3 entangles them and then destroys them. Through their prior links, this connection then entangles the states of photons 1 and 4, even though they have never interacted and may be very distant from one another. This remarkable property also has practical significance — the ability for two parties to share entanglement over long distances could have applications in secure communications and powerful distributed processing.

Even stranger things can happen. It is possible to delay the measurement on photons 2 and 3 until after photons 1 and 4 have been detected. One need not even decide whether to make that measurement until after 1 and 4 are detected. Yet the experiment seems to 'know' what you will do in advance: 1 and 4 appear entangled if a later measurement of 2 and 3 is made; they are not entangled if not.

Dance of the Photons: From Einstein to Quantum Teleportation
ANTON ZEILINGER
FSG: 2010. 320 pp. $26

It is as if photons 1 and 4 knew the future — whether or not the measurement would be made at a later time. The state of the photon not only seems to depend on the choice of measurement, but also on measurements that are yet to be made. This has implications for our ideas about reality and time, but Zeilinger reminds us that we must always make a careful accounting of the data. The reward for following Alice and Bob's reasoning as they teach us how to puzzle out these types of result is a rich understanding of entanglement beyond the simplified picture.
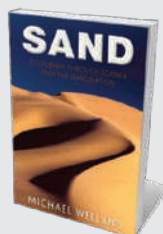
Zeilinger adds local colour throughout the book. In his tale, however, the real treasure of Vienna is not its opera, nor Ludwig Boltzmann's blackboard (which was used for the book's sketches), but a set of dark tunnels under the River Danube. These are home to a photon teleportation experiment, in which the quantum polarization state (which shows the orientation of the plane in which the light wave oscillates) of a photon on one side of
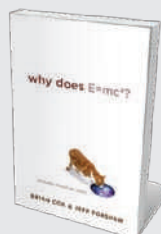

Sand: A Journey through Science and the Imagination
Michael Welland (Oxford Univ. Press, 2010; £9.99)
The world is visible in a grain of sand in geologist Michael Welland's acclaimed book. From dunes to ancient glass to electronics, he opens doors to its mysteries. "Nothing like it has been published before," wrote Andrew Robinson in his review of the hardback edition (*Nature* **460,** 798–799; 2009).

Why Does E=mc²? (And Why Should We Care?)
Brian Cox and Jeff Forshaw (Da Capo, 2010; £8.99)
Physicists Brian Cox and Jeff Forshaw provide an accessible explanation of Einstein's iconic equation. They explain the equivalence of mass and energy and look ahead to investigations of the nature of mass at the Large Hadron Collider at CERN, the particle-physics lab in Switzerland.

the Danube is instantaneously transferred to a photon on the other side. Again, the author gives the science a human face: we meet Rupert, possibly a caricature of Zeilinger's postdoc, who is condemned to the tunnels to keep the equipment running. Fortunately, Zeilinger instils him with a sense of humour.

The Vienna group's latest entanglement experiments are performed on a far larger scale — between two of the Canary Islands. A telescope with a one-metre-diameter mirror is used to catch an entangled photon that has travelled 144 kilometres through the turbulent atmosphere. Optimizing the optics, stabilizing the pointing systems and synchronizing the electronics over picoseconds make these experiments challenging, but they have enabled even more careful tests of the counter-intuitive features of quantum entanglement. By using satellites to send the quantum signals, such techniques will one day allow us to distribute entangled information between far-distant locations on Earth.

The book concludes with an outlook of where entanglement will and won't take us. Teleporting humans may be out, as we can't entangle two atom-for-atom clones of a person. But the powerful way in which quantum states carry information opens the path to quantum computing and quantum cryptography. By sharing entanglement over optical fibres (as in the Danube experiment), secret keys can be distributed over short distances. Using entanglement swapping (as in the delayed choice experiment), we might build a quantum repeater — a device for extending key distribution over much longer ranges. Using satellites, secure worldwide communication networks between classical and quantum computers will become possible.

Dance of the Photons is an enjoyable introduction to the strange world of quantum phenomena and the technologies they empower. It gives a foundation from which to ponder the nature of randomness and reality — and whether, in Vienna, the photon dance is performed to a Strauss waltz. Maybe Rupert can tell us over a lager, if he's ever allowed out of the tunnels. ■

Geoff Pryde *is associate professor of physics at Griffith University, Brisbane, Queensland 4111, Australia.*
e-mail: g.pryde@griffith.edu.au

MATHEMATICS

Deception by numbers

Jascha Hoffman reads about the rise of nonsense statistics in everything from adverts to voting.
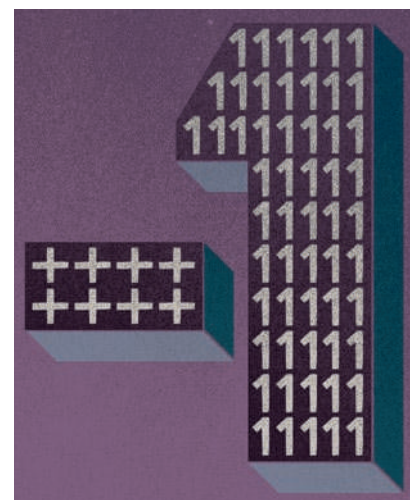

The statement, published in a newspaper, that only 0.027% of US felony convictions are wrongful is false. Based on a back-of-the-envelope calculation, it was nevertheless quoted in a court case that ended with a prisoner being sent to his death. Such bad figures are "toxic to democracy", argues science journalist and former mathematics student Charles Seife in his latest book *Proofiness*, a field guide for spotting the numeric impostors. Seife's polemic against the reporters, politicians, scientists, lawyers and bankers who spread tenacious and specious statistical claims is strident but sobering.

Seife coins the term "proofiness" to refer to the misuse of numbers, deliberate or otherwise. He dubs the simplest quantitative sins "fruit-packing". These include: "cherry-picking" the data, as he says Al Gore did when describing climate change in *An Inconvenient Truth*; "comparing apples to oranges", as economics pundits do when they neglect to adjust for price inflation; and "apple-polishing", as when advertisers use graphics to mislead.
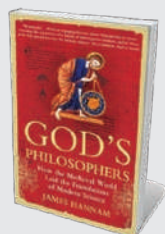
Seife finds bogus figures in every corner of public life — where there are numbers, they will be fudged. He does not spare his fellow hacks, citing the opinion poll as a method for journalists to manufacture their own stories. Surveys, no matter how large their

Proofiness: The Dark Arts of Mathematical Deception
CHARLES SEIFE
Viking: 2010. 295 pp. $25.95

sample sizes and small their margins of random error, may be skewed by slanted questions, biased samples and lying respondents, he explains.
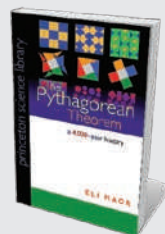
Even the simple act of counting ballots can be fraught with controversy, as in the contested Florida presidential recount in 2000. Claiming the margin of error to have been larger than the 537-vote difference between George W. Bush and Gore in that state, Seife suggests that the race should have been declared too close to call — and therefore, by Florida law, settled by the drawing of lots. He also describes economist Kenneth Arrow's impossibility theorem, which expresses how no voting system can fully capture the preferences of a group.

Seife faults some scientists, too, for over-interpreting their data and making extravagant causal inferences when the evidence is slim. This is particularly problematic in health and nutrition research, he argues,

---

God's Philosophers: How the Medieval World Laid the Foundations of Modern Science
James Hannam (Icon Books, 2010; £9.99)
Historian James Hannam debunks myths about the European 'dark ages', explaining that medieval people didn't think the world was flat. Rather, the many achievements during the period fed into the later works of Galileo and Newton.

The Pythagorean Theorem: A 4,000-Year History
Eli Maor (Princeton Univ. Press, 2010; $17.95)
Pythagoras's famous geometric theorem is central to science. Mathematics historian Eli Maor describes its origins and explains how it features in every scientific field today, pointing out that the formula was known by the Babylonians 1,000 years before Pythagoras.

casting doubt on studies alleging, for example, that an artificial sweetener causes brain cancer and that debt causes illness. He criticizes a handful of peer-reviewed articles, including some published in *Nature,* for making claims that, in his eyes, go beyond common sense. For example, Seife thinks it unlikely that wearing red helps Olympic fighters to win, offering his own analysis of results from the 2008 Beijing Olympics as proof. He dismisses other assertions, such as that wide-hipped women give birth to more sons than daughters, as mixing up cause and effect. Seife highlights how scientists can sometimes be seduced by models whose curves fit their data, attributing misguided efforts to find causal relationships to a "misfiring of our pattern-seeking behavior".

Moving on to the legal system, Seife describes how probabilities may be taken out of context in court. Statistics showing that particular crimes or events are rare have wrongly been cited as proof of innocence and guilt — delivering what Seife calls "judicial nonsense". In business, problems arise when numbers are used to under- or overstate potential dangers. Whereas the media tend to overplay risk, Seife reminds us that "underestimating risks, not exaggerating them, is where the money is". He points to prominent company directors who hid their firms' liabilities, and corporate banks that had to be bailed out by governments because of their reckless underestimation of credit risk.

Seife can overstate his case, as when he claims that proofiness is robbing us of "the democratic right to think for ourselves", oiling the "machinery of death" and "crippling our economy". He does little to explain why, given the onslaught of phony figures, many people remain susceptible to them, and he provides few practical suggestions for reducing their influence. Yet there is plenty of healthy scepticism and common sense in Seife's taxonomy of statistical malfeasance. In a world of unreliable numbers, *Proofiness* is a helpful guide. ∎

Jascha Hoffman *is a journalist based in San Francisco, California.*
e-mail: jascha@jaschahoffman.com

Crime–scene science in the dock

Two books chart the growth of forensic science from its birth to modern times, finds **Laura Spinney**.

Here are two books that span an era. Douglas Starr's *The Killer of Little Shepherds* describes the birth of modern forensic science in France in the late nineteenth century, revealing how it led to the capture of a serial killer. Michael Capuzzo's *The Murder Room* revisits cold cases from the past 50 years, just as the field of forensics is beginning to modernize and move in a new direction. Both accounts are riveting. But whereas Starr knows he is writing about a period of intellectual upheaval, Capuzzo seems impervious to the winds of change.

Starr's hero is the French physician and criminologist Alexandre Lacassagne, who established the ground rules for many forensic disciplines, from autopsy and blood-spatter analysis to toxicology and psychology. He worked in exciting times for the field. Between 1885 and the First World War, when Lacassagne's school of forensics in Lyons was influential, anthropologists Francis Galton in Britain and Juan Vucetich in Argentina were classifying fingerprint types for identification purposes, Austrian physician Karl Landsteiner discovered blood groups and, in 1897, a Parisian blaze provided the backdrop for the first identification of corpses by their teeth. The application of probability theory to the interpretation of forensic evidence in court was highlighted by the Dreyfus affair — the trial in France of artillery officer Alfred Dreyfus for treason, which hinged on the analysis of handwriting in an incriminating document.

Lacassagne brought such forensic advances to bear on the case of Joseph Vacher, a serial murderer whose

The Killer of Little Shepherds: A True Crime Story and the Birth of Forensic Science
DOUGLAS STARR
Knopf/Simon & Schuster: 2010/2011. 320 pp. $26.95/£16.99

The Murder Room: The Heirs of Sherlock Holmes Gather to Solve the World's Most Perplexing Cold Cases
MICHAEL CAPUZZO
Gotham/Michael Joseph: 2010. 448 pp/384 pp. $26/£17.99

victims included young shepherd boys out watching their flocks in rural France. Through analyses of the crime scenes and victims' bodies, the criminologist showed that Vacher's crimes were premeditated and systematic, implying that the killer was not insane. Vacher was convicted in 1898, and executed by guillotine.

Similar forensic methods are still used more than a century later. Capuzzo's heroes in *The Murder Room* are William Fleisher, a former special agent with the US Federal Bureau of Investigation, and forensic psychologist Richard Walter and forensic sculptor Frank Bender, who together founded the Vidocq Society in Philadelphia, Pennsylvania, in 1990. Taking its name from the nineteenth-century French crook-turned-crimefighter Eugène Vidocq, the non-profit, closed society brings together 150 volunteer experts to solve crimes that have gone cold. From forensic scientists to business leaders, the membership pools its knowledge once a month, over lunch, to home in on perpetrators

⊃ NATURE.COM
For a special issue focusing on science in court, see:
go.nature.com/eZ6Pwk

Origins of Human Communication
Michael Tomasello (MIT Press, 2010; £13.95)
Developmental psychologist Michael Tomasello examines the evolutionary origins of human communication. Sharing information with and helping others, he suggests, is the main purpose of speech and gesture. Such goals require the development of complex linguistic grammars.

Six-Legged Soldiers: Using Insects as Weapons of War
Jeffrey A. Lockwood (Oxford Univ. Press, 2010; £9.99)
From scorpions used by Roman armies to beetle infestations spread in the cold war, entomologist Jeffrey Lockwood reveals insects' military uses. Reviewer Kenneth J. Linthicum described it as "an excellent account" (*Nature* **456,** 36–37; 2008).

overarching theme. Rather than being about disease, his tales are more about his patients' astonishing capacities to adapt — and even thrive — in radically transformed worlds. His books resonate because they reveal as much about the force of character as they do about neurology.

The Mind's Eye also relates how an Australian psychologist named Zoltan Torey, rendered blind at 21 by a splash of acid, cultivates his photographic memory to the point that he shocks his neighbours by replacing the gutters of his house alone at night. In another chapter, Canadian novelist Howard Engel discovers that his morning *Globe and Mail* has been rendered into Cyrillic or Korean; it is his brain, of course, that has been translated by a stroke. After years of exhausting effort to engage language in new ways — composing by dictation, learning to scan words by linking adjacent letters — the novelist teaches himself to write books again.

For Sacks, disorders of vision, including his own, open a window on the brain's surprisingly active role in the authoring of experience. While under treatment for the ocular melanoma, the neurologist undertook a series of fascinating self-experiments. In one exercise, for example, he tested the limits of his brain's ability to fill in temporary gaps in his visual field caused by radiation treatment. Sacks found that repetitive patterns such as brickwork, and even clouds and trees, readily appeared to preserve the illusion of a seamless panorama around him. Faces, however, were beyond the conjuring ability of his visual cortex. "I've learned that the brain is always busy," he told me in an interview last summer.

Thankfully, Sacks's tumour has not returned, but he is still learning to cope with the aftermath, including a possibly permanent loss of three-dimensional vision — a poignant turn of events for a proud member of the New York Stereoscopic Society.

To maintain his own sense of continuity in the face of these challenges, Sacks will have to draw inspiration from the patients he has written about for 40 years. "The problems never went away," he quotes Engel as saying, "but I became cleverer at solving them". ∎

Steve Silberman *is a writer based in San Francisco, California.*
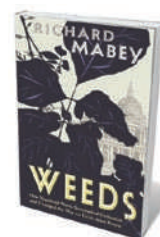*e-mail: digaman@sonic.net*

Hitchers, outcasts and wasteland beauties

Sandra Knapp revels in a portrait of weeds as resilient rebels shaped by our meddling with the wild.

Like humans, weeds are pervasive, domineering and badly behaved. But they adopt these traits only in order to reproduce. As naturalist Richard Mabey explains in *Weeds,* they are an in-your-face example of evolution by natural selection: weeding benefits weeds by allowing those that evade the hoe to produce seeds that inherit the very characteristics that allowed escape; using herbicide causes weeds to become more resistant to such poisons.

Mabey weaves social history, psychology, literature and art into his clear rendering of plant biology. Explanations of evolution sit alongside explorations of flower symbolism in Shakespeare. This blend, familiar to fans of his earlier reflections on nature in the wild, broadens the book's scope to human attitudes to plants in general.

Indeed, the concept of a weed makes sense only in relation to people — they are plants that cause us trouble by growing where we don't want them. Most of the social connotations of weeds are negative: unruly, weak or aggressive. Yet these designations are fluid. Some plants, such as St John's wort (*Hypericum perforatum*) or hemp (*Cannabis sativa*), have passed from love to hate and back again. Others, such as autumn ladies' tresses (*Spiranthes spiralis*), are a rampant but admired invader of our lawns.

Some weeds considered ubiquitous today were once rare: rosebay willowherb (*Epilobium angustifolium*), depicted among the fine flora on the ceiling of the Natural History Museum in London, was described by some nineteenth-century botanists as a woodland plant 'not often met with in the wild state'. This magenta-flowered perennial carpeted the bombed areas of 1940s London, earning it the common name of fireweed. Its

Weeds: How Vagabond Plants Gatecrashed Civilisation and Changed the Way We Think About Nature
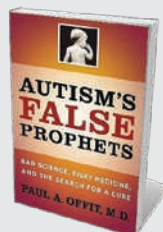RICHARD MABEY
Profile Books: 2010. 288 pp. £15.99

tiny seeds, carried on downy plumes, were dispersed by turbulence along railways; it now colonizes cities across Europe and North America. It is a good example of how weeds are a human construct, promoted by our tendency to disturb land.

Naturally invasive or easily transported species are also troublesome, particularly on islands with rare flora such as Hawaii, the Galapagos and Australia. For example, the velvet tree (*Miconia calvescens*) has taken over rainforest areas in Tahiti and is spreading on Hawaii; it chokes off native vegetation, preventing natural forest regeneration in these fragile habitats. But these plants arrived with people. *Homo sapiens* is the ultimate invasive species — coming out of Africa to colonize the globe, altering the planet beyond recognition.
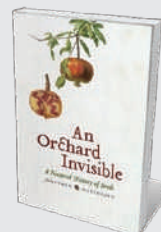
Weeds highlights our ambivalence about naturalness and artificiality. We often think of pristine nature as the landscape we, or our grandparents, grew up with. Yet nature changes all the time. In the Pleistocene, much of northern Europe was covered with ice: no plants grew. Our entire flora is invasive, but that hasn't stopped us loving it. ∎

Sandra Knapp *is a botanist at The Natural History Museum, London SW7 5BD, UK.*
e-mail: s.knapp@nhm.ac.uk

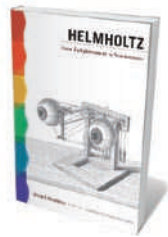HISTORY

A polymath rediscovered

George Rousseau uncovers the physiological side of Hermann von Helmholtz.

When the Harvard University psychologist Edwin Boring dedicated his classic 1942 monograph *Sensation and Perception in the History of Experimental Psychology* to Hermann von Helmholtz, many American readers wondered why. Helmholtz was a German, the Allies were rallying against the Nazi menace, and the United States had just entered the war. Few beyond professional historians of science knew about Helmholtz's work.

Boring justified his choice: "There is no one else to whom one can owe so completely the capacity to write a book about sensation and perception." Sixty years on, Helmholtz's major contributions to physiology and medicine, including his theories of visual and aural perception, have been largely eclipsed by his work in physics. In *Helmholtz*, neuroscientist Michel Meulders redresses the balance, showing that this towering figure was as influential as philosopher Immanuel Kant and as visionary as polymath Johann Wolfgang von Goethe.

Part of the reason for Helmholtz's partial invisibility today is that he wrote in German. It took decades for his work to reach the English-speaking world; his *Popular Lectures on Scientific Subjects*, delivered in the 1850s, were translated in the 1870s and 1880s, and his acoustical masterpiece, *On the Sensations of Tone as a Physiological Basis for the Theory of Music* (1863), in 1885. After this flurry of works — distributed during Helmholtz's last two decades — came tributes on his death in 1894. His Jewish student Leo Koenigsberger published a classic biography, again in German, in 1902, which was translated into English in 1906.

An extensive obituary in the 1896 *Proceedings of the Royal Society of London* portrayed Helmholtz as the most important physicist of the epoch. His work on the conservation of

Helmholtz: From Enlightenment to Neuroscience
MICHEL MEULDERS (TRANSLATED BY LAURENCE GAREY)
The MIT Press: 2010. 264 pp. $27.95, £20.95

energy that led him to formulate the first law of thermodynamics in 1847 was widely cited — electromagnetism was cutting-edge science. But interest in his physiology and medicine was lost. Helmholtz himself pursued physics more than physiology after the 1870s, and his theories of sight and sound were bitterly contested well into the twentieth century. Meulders restores Helmholtz's legacy by placing him within the history of science and by locating him as an aesthetic thinker as well as a scientist.

A welcome and surprising inclusion in the book is Helmholtz's role within the aesthetics of music. Meulders is right to retrieve this overlooked aspect — only a handful of specialized monographs have touched on it before. Helmholtz tackled the aesthetics of pitch and tone in 1857, after a century of neglect. "Music has hitherto withdrawn

MEULDERS DOESN'T SURRENDER HIS ADMIRATION — AT TIMES VERGING ON HERO WORSHIP.
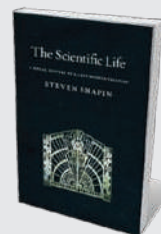
itself from scientific treatment, more than any other art," he wrote. Poetry, painting and sculpture borrow from the world of experience, he explained, but music seems to "reject all anatomization of pleasurable sensations".

Helmholtz developed a 'resonator' device, a pierced sphere of glass or brass with a narrow neck, to demonstrate musical pitch and tonal

colour. His view was that music depends on human experience and on the physiology of the senses for its effects. Helmholtz's physiological theory of music had a lasting impact on the composers Alexander Scriabin and Nikolai Rimsky-Korsakov, and on many twentieth-century academic musicologists.

Meulders brings in other German intellectuals on whose work Helmholtz built. For example, he analyses the theory of physics and physiology of colours published by Goethe as *Zur Farbenlehre* (Colour Theory) in 1810. Yet Goethe does not come to life in the book in the same way as Helmholtz's teacher Johannes Müller, portrayed as a genius who overcame insomnia and depression to hew a science of physiology.

Müller demonstrated in his famous Berlin laboratory that "the results of all physiological research must be, in the end, psychological in nature". Small wonder, then, that he assigned to his protégé Helmholtz a doctoral thesis topic in the 1830s based on invertebrates in Müller's own collection, which was eventually published as *Nerve Fibres Arising from the Ganglion Cells Discovered in 1836*. In this, Helmholtz built on the ideas of his teacher to bring together physiology and psychology.

Yet curiously, Meulders writes, Helmholtz never referred to the brain. My main reservation is that the book does not unpack this statement. Helmholtz consistently ignored anatomical data on the nervous system, and probably mistrusted the concept, popular at the time, that anatomical and psychological processes were identical. Thus he did not link the psychology of perception with the physical brain, and bought into an older theory of mind, with the soul as the arbiter of the senses. Helmholtz's defiance of nineteenth-century natural philosophy through his enduring omission of the brain is strange, and I hope another author will pursue it.

Imperial Nature: Joseph Hooker and the Practices of Victorian Science
Jim Endersby (Univ. Chicago Press, 2010; $25)
Botanist Joseph Hooker became one of the first professional scientists when research began to be funded by governments. "A refreshing record of how scientists worked during this transition," wrote Sandra Knapp (*Nature* **453**, 721; 2008).

The Scientific Life
Steven Shapin (Univ. Chicago Press, 2010; $20)
Historian Steven Shapin shatters myths about the divide between pure and commercial science by arguing that moral values are as abundant in industry as in academia. Reviewer Jerome Ravetz described it as "required reading for all scientists" (*Nature* **457**, 662–663; 2009).

Meulders concludes his book with three incisive chapters on the aesthetics of music. In one he deals with the Pythagorean legacy, especially the idea that mathematical relationships were the basis of harmony and tone. In the second he considers 'the musical ear', demonstrating that findings in auricular physiology, particularly Italian anatomist Alfonso Corti's discovery in 1851 of fibres that function as acoustical sensor cells in the cochlea, had complicated the aesthetics of sound.

This chapter is a triumph of compression of a vast province of physiology and aesthetics into a few pages. Surveying the musicological terrain from the argument between Jean-Philippe Rameau and Jean le Rond d'Alembert to Johann Sebastian Bach and Andreas Werckmeister, and on to Mozart and Mendelssohn, Meulders pauses to explain how Helmholtz the empiricist understood music theory and aesthetics as a grand unifier. Musical sounds, he thought, can only be understood as great art by combining anatomy, physiology, philosophy and psychology. The third of these chapters meditates on Helmholtz's nostalgia, intuition and memory — an odd amalgam, the breadth of which adds to Meulders's claim for Helmholtz's genius.

NATURE.COM
For more on German science history, see:
go.nature.com/R5K7Qw

Meulders stitches together the thoughts of a lifetime into his slim book. He doesn't surrender his admiration — at times verging on hero worship — despite the occasional critique. The approach is hit-and-miss and does not amount to the much-desired extended interpretation unifying Helmholtz's physiology and aesthetics, but it is a brave start.

Meulders sums up his subject thus: "With his will to unify so many different scientific disciplines in a coherent entity, he proved once again his veritable gluttony for science and knowledge." Some may find Meulders equally gluttonous, but his book demonstrates that Helmholtz was indeed a polymath *par excellence*. ∎

George Rousseau is a professor of history and co-director of the Centre for the History of Childhood, University of Oxford, Oxford, OX1 4AU, UK. He is author of Nervous Acts.
e-mail: george.rousseau@magd.ox.ac.uk

ECOLOGY

Conservation thriller earns its stripes

A travelogue about tiger poaching in Russia's far east opens up a new genre, discovers **Geoff Marsh**.

Yuri Trush steadily points his camera at the stubs of bone protruding from a pair of thin rubber boots lying in the blood-speckled snow. As the leader of an Inspection Tiger anti-poaching unit, his job now is to piece together the details of Vladimir Markov's run-in with the tiger. Judging by the whimpering of Trush's dog, the big cat in question remains close by, among the trees.

Inspection Tiger is a government agency that was set up to combat poaching in Primorskiy Kray (or Primorye) — an area the size of Washington state in the far east of Russia, bordered by China and North Korea. Trush's team travels in a decommissioned army truck, armed with knives, pistols and semiautomatic rifles. Their mission is to intercept poachers and to resolve the locals' conflicts with the largest cats in the world.
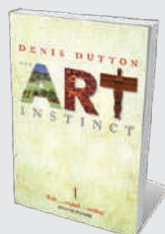
In *The Tiger*, author John Vaillant relates his travels across the region while investigating the pressures on tiger conservation. His vivid portrayal of Primorye reveals a unique ecosystem at the crossroads of four distinct biomes: the Siberian taiga forest, the steppes of Mongolia, the subtropics of Manchuria and the boreal forest of the far north. A peculiar mix of hardy alpine

The Tiger: A True Story of Vengeance and Survival
JOHN VAILLANT
Sceptre/Alfred Knopf: 2010. 352 pp. £18.99/$26.95

The Art Instinct
Denis Dutton (Oxford Univ. Press, 2010; £9.99)
Art appreciation has an evolutionary basis, according to philosopher Denis Dutton. The basic elements of aesthetic taste are similar across cultures and are part of our evolutionary heritage rather than being socially constructed, he claims provocatively.

Pink Brain, Blue Brain
Lise Eliot (OneWorld, 2010; £12.99)
Neuroscientist Lise Eliot marshals the latest evidence to show that social pressures are the main cause of behaviour differences between boys and girls. Although small gender variations are apparent at birth, they grow as our plastic brains quickly become modified by experiences.

and lush tropical plants shelter an equally varied assortment of animals — timber wolves compete with leopards for fanged musk deer.

The Amur (or Siberian) tiger is one of nine recognized subspecies, three of which have gone extinct in the past century. Their numbers in Russia have declined severely during this time. The period 1992–94 alone saw one-quarter of the country's wild tiger population killed and sold, mostly to China for use in traditional medicine. Last year, the international Siberian Tiger Monitoring Programme reported a significant drop in numbers in the past decade; now, probably fewer than 400 tigers remain in the Russian far east. Poaching is thought to be the main factor in their decline.

The tension between humans and tigers first arose from a shared appetite for meat and large territories, says Vaillant. Add to this the poverty of many of the inhabitants of Primorye after perestroika in the late 1980s and the temptations of a lucrative black market for tiger parts, and cases such as Markov's become inevitable. People must poach or starve.

Vaillant weaves his story using an evolutionary and cultural context. Our relationship with big cats began with us scavenging their kills, he suggests. Predation was of secondary concern, with humans taking the risk of being attacked in order to scavenge, and both species largely leaving each other alone. This evolutionary treaty to do no harm is reflected in the behaviours of

LOCALS BELIEVE THAT THE TIGER WILL BE
PURPOSEFULLY VENGEFUL
AGAINST POACHERS.

the native hunters in Russia's far east, and in the relationship of Kalahari bushmen with lions: both groups avoid confrontation with the cats, and are able to live safely alongside them.

When a Primorye poacher goes against this treaty, the locals believe that the tiger will be purposefully vengeful. Markov reportedly shot at the tiger that killed him days before the attack; the tiger then waited at his cabin for him to return. Although clearly anthropomorphized, this theory of feline vendetta is a haunting notion.

The Tiger does more than paint a gloomy picture of the Amur tigers' demise in northeast Asia. Vaillant points out that the animal's fate is entirely in our hands. Its conservation represents more than just the survival of this charismatic predator: because it is a keystone species, an environment in which a tiger thrives is necessarily a healthy one. The very presence of tigers at the top of an ecosystem confirms that it is intact. Vaillant describes the tiger as "an enormous canary in the biological coal mine".
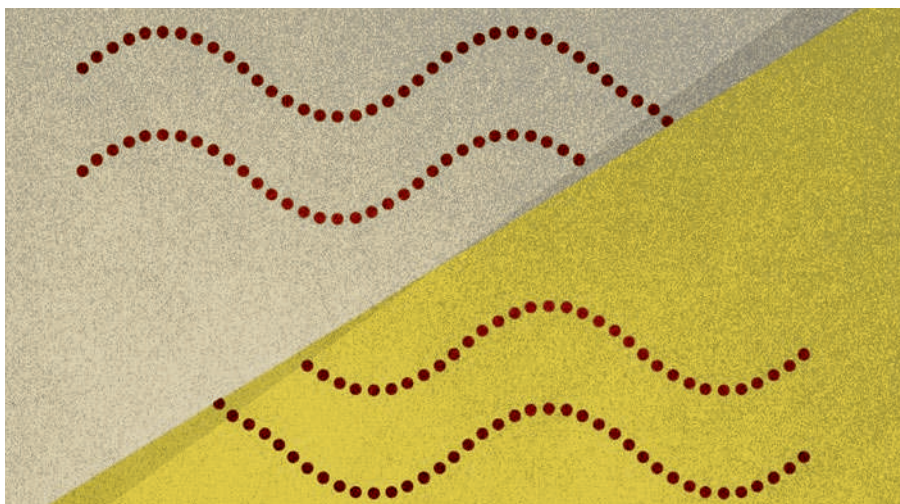
Heroes such as Trush and his team are as endangered as the tigers they protect, owing to severe cuts in staff and funding. Restoring such agencies, Vaillant says, will be key to the survival of the Amur tiger and its prey.

This epic story helps to raise awareness of conservation issues in the Russian far east, yet its reach is greater: actor Brad Pitt and film director Darren Aronofsky are currently adapting *The Tiger* for the big screen. This new genre of conservation thriller could be a powerful way of generating interest in the plight of species that are on the brink of extinction. ∎

Geoff Marsh *is a former ecologist who is now a multimedia producer at* Nature.

QUANTUM PHYSICS

Tripping the light fantastic

Geoff Pryde on the weird world of quantum entanglement.

The only way to understand the quantum world is to measure it. This empirical view is dear to the heart of Anton Zeilinger, now at the University of Vienna, a leading figure in quantum physics through his work on correlated photons. In *Dance of the Photons,* he explores the phenomenon of quantum entanglement, the quantum correlations in the properties of particles.

When two photons are made to interact, they share their quantum information and become 'entangled'. If one travels off, it retains knowledge about its counterpart. So measuring one can determine the state of the other, even if they are far apart. Albert Einstein was worried by such reasoning: instant messaging between entangled particles contradicted his theory of relativity, which stated that signals cannot travel faster than the speed of light, unless you allow the crazy idea that particles do not have real properties independent of measurement. Quantum mechanics, he

decided, was not up to explaining the world.

Zeilinger explains that Einstein was wrong. Experiments in the 1980s and 1990s proved the weird predictions of quantum entanglement to be true. Putting the reader in the role of discoverer, he describes these tests through the eyes of fictional students Alice and Bob, namesakes of the characters regularly put to work in explaining quantum physics. Examining the philosophical and technological implications of spooky quantum phenomena, he points to big issues that demand further thought — the inherent randomness of quantum physics and the role of the observer in determining a quantum particle's reality.
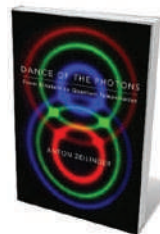
As well as giving an overview of other work, Zeilinger relates in detail his own group's research. For instance, he describes a 'delayed choice entanglement swapping' experiment he has carried out using four photons (1, 2, 3, 4). Two pairs share prior information: photons 1 and 2 are entangled, photons 3 and 4

are also entangled, but there is no correlation between those pairs. Making a particular type of quantum measurement — known as a Bell measurement — jointly on photons 2 and 3 entangles them and then destroys them. Through their prior links, this connection then entangles the states of photons 1 and 4, even though they have never interacted and may be very distant from one another. This remarkable property also has practical significance — the ability for two parties to share entanglement over long distances could have applications in secure communications and powerful distributed processing.

Even stranger things can happen. It is possible to delay the measurement on photons 2 and 3 until after photons 1 and 4 have been detected. One need not even decide whether to make that measurement until after 1 and 4 are detected. Yet the experiment seems to 'know' what you will do in advance: 1 and 4 appear entangled if a later measurement of 2 and 3 is made; they are not entangled if not.

Dance of the Photons: From Einstein to Quantum Teleportation
ANTON ZEILINGER
FSG: 2010. 320 pp. $26

It is as if photons 1 and 4 knew the future — whether or not the measurement would be made at a later time. The state of the photon not only seems to depend on the choice of measurement, but also on measurements that are yet to be made. This has implications for our ideas about reality and time, but Zeilinger reminds us that we must always make a careful accounting of the data. The reward for following Alice and Bob's reasoning as they teach us how to puzzle out these types of result is a rich understanding of entanglement beyond the simplified picture.
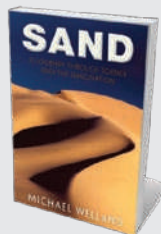
Zeilinger adds local colour throughout the book. In his tale, however, the real treasure of Vienna is not its opera, nor Ludwig Boltzmann's blackboard (which was used for the book's sketches), but a set of dark tunnels under the River Danube. These are home to a photon teleportation experiment, in which the quantum polarization state (which shows the orientation of the plane in which the light wave oscillates) of a photon on one side of
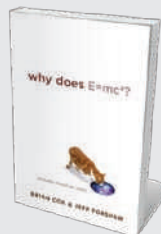

Sand: A Journey through Science and the Imagination
Michael Welland (Oxford Univ. Press, 2010; £9.99)
The world is visible in a grain of sand in geologist Michael Welland's acclaimed book. From dunes to ancient glass to electronics, he opens doors to its mysteries. "Nothing like it has been published before," wrote Andrew Robinson in his review of the hardback edition (*Nature* **460,** 798–799; 2009).

Why Does E=mc²? (And Why Should We Care?)
Brian Cox and Jeff Forshaw (Da Capo, 2010; £8.99)
Physicists Brian Cox and Jeff Forshaw provide an accessible explanation of Einstein's iconic equation. They explain the equivalence of mass and energy and look ahead to investigations of the nature of mass at the Large Hadron Collider at CERN, the particle-physics lab in Switzerland.

the Danube is instantaneously transferred to a photon on the other side. Again, the author gives the science a human face: we meet Rupert, possibly a caricature of Zeilinger's postdoc, who is condemned to the tunnels to keep the equipment running. Fortunately, Zeilinger instils him with a sense of humour.

The Vienna group's latest entanglement experiments are performed on a far larger scale — between two of the Canary Islands. A telescope with a one-metre-diameter mirror is used to catch an entangled photon that has travelled 144 kilometres through the turbulent atmosphere. Optimizing the optics, stabilizing the pointing systems and synchronizing the electronics over picoseconds make these experiments challenging, but they have enabled even more careful tests of the counter-intuitive features of quantum entanglement. By using satellites to send the quantum signals, such techniques will one day allow us to distribute entangled information between far-distant locations on Earth.

The book concludes with an outlook of where entanglement will and won't take us. Teleporting humans may be out, as we can't entangle two atom-for-atom clones of a person. But the powerful way in which quantum states carry information opens the path to quantum computing and quantum cryptography. By sharing entanglement over optical fibres (as in the Danube experiment), secret keys can be distributed over short distances. Using entanglement swapping (as in the delayed choice experiment), we might build a quantum repeater — a device for extending key distribution over much longer ranges. Using satellites, secure worldwide communication networks between classical and quantum computers will become possible.

Dance of the Photons is an enjoyable introduction to the strange world of quantum phenomena and the technologies they empower. It gives a foundation from which to ponder the nature of randomness and reality — and whether, in Vienna, the photon dance is performed to a Strauss waltz. Maybe Rupert can tell us over a lager, if he's ever allowed out of the tunnels. ∎

Geoff Pryde *is associate professor of physics at Griffith University, Brisbane, Queensland 4111, Australia.*
e-mail: g.pryde@griffith.edu.au

Deception by numbers

Jascha Hoffman reads about the rise of nonsense statistics in everything from adverts to voting.
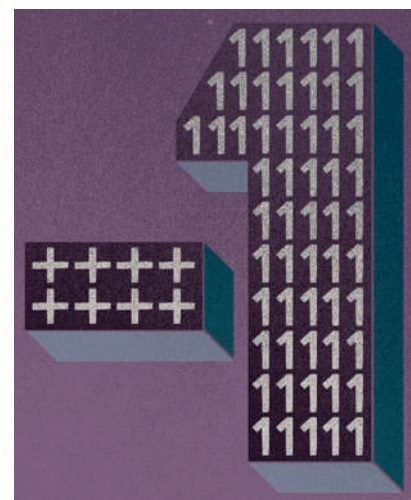
The statement, published in a newspaper, that only 0.027% of US felony convictions are wrongful is false. Based on a back-of-the-envelope calculation, it was nevertheless quoted in a court case that ended with a prisoner being sent to his death. Such bad figures are "toxic to democracy", argues science journalist and former mathematics student Charles Seife in his latest book *Proofiness*, a field guide for spotting the numeric impostors. Seife's polemic against the reporters, politicians, scientists, lawyers and bankers who spread tenacious and specious statistical claims is strident but sobering.

Seife coins the term "proofiness" to refer to the misuse of numbers, deliberate or otherwise. He dubs the simplest quantitative sins "fruit-packing". These include: "cherry-picking" the data, as he says Al Gore did when describing climate change in *An Inconvenient Truth*; "comparing apples to oranges", as economics pundits do when they neglect to adjust for price inflation; and "apple-polishing", as when advertisers use graphics to mislead.
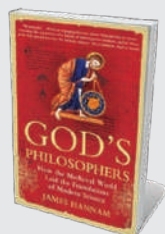
Seife finds bogus figures in every corner of public life — where there are numbers, they will be fudged. He does not spare his fellow hacks, citing the opinion poll as a method for journalists to manufacture their own stories. Surveys, no matter how large their sample sizes and small their margins of random error, may be skewed by slanted questions, biased samples and lying respondents, he explains.
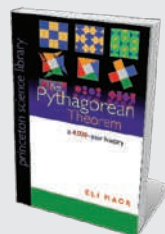
Even the simple act of counting ballots can be fraught with controversy, as in the contested Florida presidential recount in 2000. Claiming the margin of error to have been larger than the 537-vote difference between George W. Bush and Gore in that state, Seife suggests that the race should have been declared too close to call — and therefore, by Florida law, settled by the drawing of lots. He also describes economist Kenneth Arrow's impossibility theorem, which expresses how no voting system can fully capture the preferences of a group.

Seife faults some scientists, too, for over-interpreting their data and making extravagant causal inferences when the evidence is slim. This is particularly problematic in health and nutrition research, he argues,

Proofiness: The Dark Arts of Mathematical Deception
CHARLES SEIFE
Viking: 2010.
295 pp. $25.95

God's Philosophers: How the Medieval World Laid the Foundations of Modern Science
James Hannam (Icon Books, 2010; £9.99)
Historian James Hannam debunks myths about the European 'dark ages', explaining that medieval people didn't think the world was flat. Rather, the many achievements during the period fed into the later works of Galileo and Newton.

The Pythagorean Theorem: A 4,000-Year History
Eli Maor (Princeton Univ. Press, 2010; $17.95)
Pythagoras's famous geometric theorem is central to science. Mathematics historian Eli Maor describes its origins and explains how it features in every scientific field today, pointing out that the formula was known by the Babylonians 1,000 years before Pythagoras.

casting doubt on studies alleging, for example, that an artificial sweetener causes brain cancer and that debt causes illness. He criticizes a handful of peer-reviewed articles, including some published in *Nature,* for making claims that, in his eyes, go beyond common sense. For example, Seife thinks it unlikely that wearing red helps Olympic fighters to win, offering his own analysis of results from the 2008 Beijing Olympics as proof. He dismisses other assertions, such as that wide-hipped women give birth to more sons than daughters, as mixing up cause and effect. Seife highlights how scientists can sometimes be seduced by models whose curves fit their data, attributing misguided efforts to find causal relationships to a "misfiring of our pattern-seeking behavior".

Moving on to the legal system, Seife describes how probabilities may be taken out of context in court. Statistics showing that particular crimes or events are rare have wrongly been cited as proof of innocence and guilt — delivering what Seife calls "judicial nonsense". In business, problems arise when numbers are used to under- or overstate potential dangers. Whereas the media tend to overplay risk, Seife reminds us that "underestimating risks, not exaggerating them, is where the money is". He points to prominent company directors who hid their firms' liabilities, and corporate banks that had to be bailed out by governments because of their reckless underestimation of credit risk.

Seife can overstate his case, as when he claims that proofiness is robbing us of "the democratic right to think for ourselves", oiling the "machinery of death" and "crippling our economy". He does little to explain why, given the onslaught of phony figures, many people remain susceptible to them, and he provides few practical suggestions for reducing their influence. Yet there is plenty of healthy scepticism and common sense in Seife's taxonomy of statistical malfeasance. In a world of unreliable numbers, *Proofiness* is a helpful guide. ∎

Jascha Hoffman *is a journalist based in San Francisco, California.*
e-mail: jascha@jaschahoffman.com

Crime–scene science in the dock

Two books chart the growth of forensic science from its birth to modern times, finds **Laura Spinney**.

Here are two books that span an era. Douglas Starr's *The Killer of Little Shepherds* describes the birth of modern forensic science in France in the late nineteenth century, revealing how it led to the capture of a serial killer. Michael Capuzzo's *The Murder Room* revisits cold cases from the past 50 years, just as the field of forensics is beginning to modernize and move in a new direction. Both accounts are riveting. But whereas Starr knows he is writing about a period of intellectual upheaval, Capuzzo seems impervious to the winds of change.

Starr's hero is the French physician and criminologist Alexandre Lacassagne, who established the ground rules for many forensic disciplines, from autopsy and blood-spatter analysis to toxicology and psychology. He worked in exciting times for the field. Between 1885 and the First World War, when Lacassagne's school of forensics in Lyons was influential, anthropologists Francis Galton in Britain and Juan Vucetich in Argentina were classifying fingerprint types for identification purposes, Austrian physician Karl Landsteiner discovered blood groups and, in 1897, a Parisian blaze provided the backdrop for the first identification of corpses by their teeth. The application of probability theory to the interpretation of forensic evidence in court was highlighted by the Dreyfus affair — the trial in France of artillery officer Alfred Dreyfus for treason, which hinged on the analysis of handwriting in an incriminating document.

Lacassagne brought such forensic advances to bear on the case of Joseph Vacher, a serial murderer whose

The Killer of Little Shepherds: A True Crime Story and the Birth of Forensic Science
DOUGLAS STARR
Knopf/Simon & Schuster: 2010/2011. 320 pp. $26.95/£16.99

The Murder Room: The Heirs of Sherlock Holmes Gather to Solve the World's Most Perplexing Cold Cases
MICHAEL CAPUZZO
Gotham/Michael Joseph: 2010. 448 pp/384 pp. $26/£17.99

victims included young shepherd boys out watching their flocks in rural France. Through analyses of the crime scenes and victims' bodies, the criminologist showed that Vacher's crimes were premeditated and systematic, implying that the killer was not insane. Vacher was convicted in 1898, and executed by guillotine.

Similar forensic methods are still used more than a century later. Capuzzo's heroes in *The Murder Room* are William Fleisher, a former special agent with the US Federal Bureau of Investigation, and forensic psychologist Richard Walter and forensic sculptor Frank Bender, who together founded the Vidocq Society in Philadelphia, Pennsylvania, in 1990. Taking its name from the nineteenth-century French crook-turned-crimefighter Eugène Vidocq, the non-profit, closed society brings together 150 volunteer experts to solve crimes that have gone cold. From forensic scientists to business leaders, the membership pools its knowledge once a month, over lunch, to home in on perpetrators
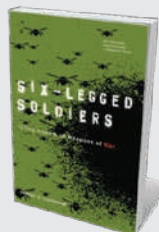
⟳ **NATURE.COM**
For a special issue focusing on science in court, see:
go.nature.com/eZ6Pwk

Origins of Human Communication
Michael Tomasello (MIT Press, 2010; £13.95)
Developmental psychologist Michael Tomasello examines the evolutionary origins of human communication. Sharing information with and helping others, he suggests, is the main purpose of speech and gesture. Such goals require the development of complex linguistic grammars.

Six-Legged Soldiers: Using Insects as Weapons of War
Jeffrey A. Lockwood (Oxford Univ. Press, 2010; £9.99)
From scorpions used by Roman armies to beetle infestations spread in the cold war, entomologist Jeffrey Lockwood reveals insects' military uses. Reviewer Kenneth J. Linthicum described it as "an excellent account" (*Nature* **456,** 36–37; 2008).

and to avenge forgotten victims. They do so because they value justice, and because they enjoy the chase.

Capuzzo describes the Vidocq Society's successes, including the identification of John List, who murdered five members of his family in 1971 and remained a fugitive for some 17 years. But what is striking about *The Murder Room* is that — with the notable exception of DNA profiling — the twentieth century added little to the nineteenth-century foundations of forensics. If Lacassagne attended a Vidocq Society lunch today, most of the techniques discussed would be familiar to him. Two modern techniques that he would not recognize — the lie detector and criminal profiling — are popular with law enforcers, although their efficacy has never been clearly demonstrated.

Together, these two books give the impression that the late nineteenth century was a golden era for forensic science and that the field has been treading water since then. Yet it is currently experiencing a crisis, which has been brewing since the advent of DNA profiling in the 1980s. Because DNA analysis had already been thoroughly validated in the academic context, its introduction raised the scientific bar for all forensic techniques — and many of them have been found wanting.
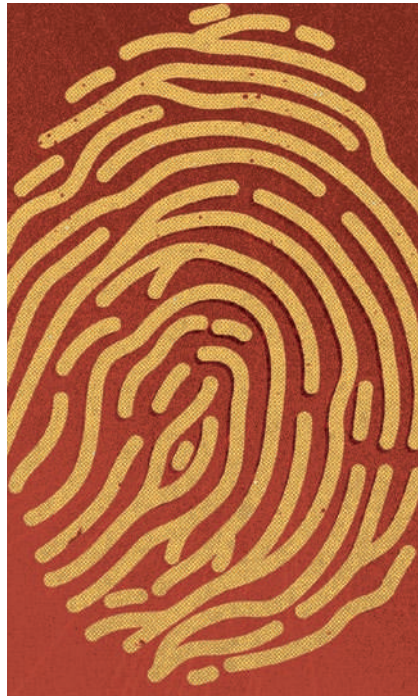
In February 2009, the US National Research Council (NRC) published a highly critical report that challenged forensic science to demonstrate its scientific credentials. The report pointed out, for example, that fingerprint analysts' long-standing claims of zero error rates were not scientifically plausible. Almost all of the techniques in use in forensic labs today — from ballistics to analyses of handwriting, shoe prints and blood patterns — came in for criticism. The NRC's message to forensic science was clear: either drag yourself out of the nineteenth century, or the police and the courts will sideline you. Yet the problem is not only in the United States — modernization of the whole field, along with the laborious empirical testing which that will entail, seems inevitable worldwide.

Capuzzo's book may unwittingly describe the end of an era. Because members of the Vidocq Society rely on law enforcers to feed them cold cases, they too will have to respond to the challenge of modernization. As nineteenth-century French forensics pioneer Alphonse Bertillon discovered to his cost in seeking the truth — his reputation was destroyed after he failed to apply probability theory correctly and wrongly attributed that damning scrawl to Dreyfus — the road to hell is paved with good intentions. It is better, in the end, to have good tools. ∎

Laura Spinney *is a writer based in Lausanne, Switzerland.*
e-mail: lfspinney@googlemail.com

CORRESPONDENCE

Marine stewardship: catalysing change

Criticisms that the Marine Stewardship Council (MSC) programme is not delivering on its promise (*Nature* **467**, 28–29 and 531; 2010) are misplaced. After ten years of operation, the MSC certification programme is helping to generate real benefits for the marine environment, including increased stock health, reduced by-catch, established no-take zones, reduced impact on marine habitats and improved scientific understanding through research.

For example, as a condition of certification, the South African hake fishery implemented measures that have reduced by-catch of seabirds from 18,000 per year to less than 200. The Dutch Ekofish North Sea plaice fishery has established a voluntary agreement with the World Wide Fund for Nature (WWF) and the North Sea Foundation to minimize impact by closing selected areas to fishing. In British Columbia, certification of the Nass River sockeye-salmon fishery is contingent on implementation of an effective recovery plan for chum salmon stocks. The Norwegian biotech firm Aker BioMarine has undertaken new research and surveys to ensure even better future management of the Antarctic krill resource (in a fishery that, in total, takes less than 1% of the available biomass). There is a proven ecological case for credible third-party certification.

The MSC has committed considerable resources to its developing-world programme, in particular by developing the Risk-Based Framework methodology for assessing data-poor fisheries. This is being used in assessments of pole and line and hand-line tuna fisheries in the Maldives, and in the Sian Ka'an and Banco Chinchorro lobster fishery in Mexico. An increasing number of fisheries in Africa, Asia and small island states in the Pacific Ocean are all engaged at various stages of the independent assessment process and we expect this trend to continue. We have expert developing-world representatives on our Technical Advisory Board and Stakeholder Council.

We — along with many scientists, experts, partners and stakeholders worldwide — have confidence in the rigour of our standard and methodology. The MSC is helping to transform the way the oceans are fished. More than 90% of the world's fisheries are not MSC certified: engaging those fisheries to achieve and establish their sustainability is the challenge that faces us all.
Rupert Howes *Marine Stewardship Council, UK, rupert.howes@msc.org*

Pakistan: why the reforms need work

I agree that investing in Pakistan's higher education will have a broad impact on development (*Nature* **467**, 367; 2010). But at policy level, some things are different from the situation described by you and the Higher Education Commission overseeing this reform process.

The commission must prioritize according to the country's needs. For example, we badly need social scientists (economists, sociologists and anthropologists) to help to set goals of human development and social welfare.

The commission is unrealistic in suggesting that producing more PhDs locally and from advanced industrial countries will boost the knowledge economy. Establishing new universities in remote districts is unlikely to attract more foreign graduates and invitee professors, who will continue to favour the metropolitan universities because of their better infrastructure.

The reform process is being partly funded by foreign partners, but it is not clear how much longer this can be sustained. And Pakistan's low tax-to-GDP ratio, coupled with burgeoning corruption (tax theft), won't help to increase local funding for higher education.
Faisal Abbas *University of Bonn, Germany, fabbas@uni-bonn.de*

Pakistan: the brain drain dilemma

In your assessment of the bleak state of academic and scientific research in Pakistan (*Nature* **467**, 378–379; 2010), you do not mention the country's 'brain drain' problem.

A nation's research achievements depend mainly on the experience and expertise of its available researchers. But the current trend for Pakistan's new PhDs is to pursue their postdoctoral training abroad and eventually to take up employment there. Few of these well-trained researchers return home, discouraged by factors such as corruption, political instability, lack of governmental initiative and inadequate health-care and social-security benefits.

In the absence of resident high-calibre scientists, even adequate funding will make little or no difference to the existing system.
Yajnavalka Banerjee *Sultan Qaboos University, Oman, yaj@squ.edu.om*

Safaris can help conservation

Conservation doesn't always alleviate poverty, and commercial ecotourism doesn't always protect biodiversity (*Nature* **467**, 264–265; 2010) — but both succeed often enough to be worth doing.

A few tourism enterprises have made globally significant contributions to conservation. The safari company &Beyond, for example, protects 2% of the world's black rhinos and 1% of white rhinos on two of its 50 properties, as well as 4% of the Aders' duiker (*Cephalophus adersi*) antelope population and 10% of suni antelopes (*Neotragus moschatus*) on two others.

In addition, Wilderness Safaris protects 8% of the world's remaining population of an endangered bird, the Seychelles white-eye (*Zosterops modestus*) on one of the company's 60 properties. For further details, see go.nature.com/g8Z4Pj.
Ralf Buckley *Griffith University, Queensland, Australia, r.buckley@griffith.edu.au*

Fate of 'retired' research chimps

Your News story on the return of a colony of elderly research chimpanzees to the lab (*Nature* **467**, 507–508; 2010) inadvertently misrepresents my position on an important and sensitive issue.

Like many others on the sidelines of this acrimonious debate, I see a middle path that seems reasonable. Given that chimpanzees in captivity cannot be returned to the wild, these individuals should be studied with care. This means following similar conditions and principles to those used for research on human subjects who are incapable of giving informed consent. Such studies would be of great benefit to chimpanzees and to humans.

I do not understand the call for a complete ban on all research on captive chimpanzees. Would anyone support a complete ban on all research on humans? Such a ban would be bad for both species.
Ajit Varki *University of California, San Diego, USA, a1varki@ucsd.edu*

Georges Charpak
(1924–2010)

Physicist who transformed the measurement of high–energy particles.

Physicist and campaigner, Georges Charpak has left an enduring mark on science, technology and education. His invention of a type of particle detector — the multiwire proportional chamber — revolutionized the collection of data from high-energy physics experiments. The device allowed physicists to detect new particles and so test fundamental theories about the nature of matter. Modern variants of the detector are still used in high-energy particle accelerators.

Charpak, who died on 29 September, was born in eastern Poland to a poor Jewish family. When he was seven, the family moved to Paris, lured by France's healthier economy. After France surrendered to Germany in 1940, Charpak refused to wear the yellow Star of David, required by Nazi authorities to identify Jews, and he became active in the French Resistance. He was imprisoned by the Vichy government of France in 1943 before being transferred to the Dachau concentration camp in 1944. He survived because the German guards did not realize that their political prisoner was actually Jewish.

After the war, Charpak became a French citizen. In 1954, he received his doctorate in nuclear physics from the Collège de France in Paris where he studied in the laboratory of the Nobel laureate Frédéric Joliot-Curie. He devoted his early career to nuclear physics before switching to high-energy particle physics under the guidance of Leon Lederman at CERN, Europe's particle-physics laboratory near Geneva, Switzerland.

In 1968, while still at CERN, but by then leading a small research group of his own, Charpak developed the multiwire proportional chamber.

When high-energy collisions occur between particles in an accelerator, they generate new charged particles that ionize the detector gas, leaving behind a trail of electrons and positive ions. Early detectors, such as the bubble chamber, worked by taking photographs of the tracks left by these charged particles moving through a medium (often liquid hydrogen in the case of the bubble chamber). Yet such devices could generate only a few photographs per second. Charpak's multiwire chamber was a gas-filled box containing a large number of parallel detector wires, each connected to individual amplifiers. It recorded the electronic pulses resulting from charged particles passing through the gas. These signals could be fed directly into a computer, increasing the detection rate of particles a thousand-fold.

Others had attempted to invent a similar device but without success — largely because it was unclear what was producing the electronic signals in the wires. Working with similar detectors in the Collège de France, Charpak realized that the electronic pulses were produced not by drifting electrons but rather by positive ions, which induced pulses of opposite polarity in the wires. This discovery led him to make large flat detectors containing several wires. Charpak's insight meant the position of a particle could be tracked with unprecedented precision.

BELATED PRIZE

The speed and precision of the multiwire chamber and its descendants, the drift chamber and the time projection chamber, have allowed physicists to operate experiments at much higher particle collision rates and so test new theories about the nature of matter. In recognition of the importance of his work on this and other detectors, Charpak was awarded the Nobel Prize in Physics in 1992.

Years before Charpak received his prize, Nobels were awarded to physicists Samuel Ting and Carlo Rubbia, for their discoveries of the J/ψ particle, and the heavy W and Z particles, respectively. Both scientists made their findings using multiwire chambers. Indeed, many of the new particles discovered in the past few decades have used detectors developed or greatly improved by Charpak and his team.

From the moment Charpak began working on detectors, he was interested in the their medical applications. Although a long-time proponent of nuclear energy, he was horrified by the radiation doses that children were exposed to during routine medical X-rays. He helped co-found several companies that applied his multiwire detectors to medical imaging, to reduce the exposure of patients to radioactive tracers. He also worked closely with surgeons and radiologists to bring these techniques to clinical settings.

Influenced by his experiences in wartime Europe, Charpak's deep concern for social issues led him to apply his knowledge to education. In 1996 he created *La main à la pâte*, an organization that introduced hands-on science education in primary schools in France. He got the idea from his old colleague Lederman, who had introduced a similar physics education programme in Chicago a few years earlier. *La main à la pâte* has now spread from France to other countries.

In 2001, he and nuclear physicist Richard Garwin argued in their book *Megawatts and Megatons: a Turning Point in the Nuclear Age?* that nuclear energy could provide an assured, economically feasible and environmentally sustainable supply of energy without driving weapons proliferation. Three years later, Charpak and Henri Broch co-authored *Debunked! ESP, Telekinesis, and other Pseudoscience*, in which they dismantled claims from parapsychology and astrology.

Georges disliked the new generation of digital detector devices. When he came to visit my laboratory at Saclay, he'd use an old instrument that we kept especially for him. He was excited, however, by a new radon detector he was developing. Indeed, he believed that this detector would have enough industrial success to allow him to "buy a new pair of shoes". Georges will be remembered as a humanist, an enthusiast, an optimist — and someone always open to new ideas. ■

Ioannis Giomataris *is research director at CEA-Saclay 91191 Gif-sur-Yvette Cedex. France.*
e-mail: ioanis.giomataris@cern.ch

A. BRUCELLE/SYGMA/CORBIS

NEWS & VIEWS

In search of rare human variants

The 1000 Genomes Project has completed its pilot phase, sequencing the whole genomes of 179 individuals and characterizing all the protein–coding sequences of many others. Welcome to the third phase of human genomics. SEE ARTICLE P.1061

RASMUS NIELSEN

The goal of the 1000 Genomes Project[1] is to find most of the variants in the human genome that have a frequency of at least 1% in the populations studied. The consortium of researchers participating in the project now reports the results of its pilot phase (page 1061 of this issue[2]).

But first let's take a step back. A decade ago, the reference copy of the human genome was sequenced[3,4]. Although that project is undoubtedly one of the greatest scientific achievements of our time, its potential societal impact will be fully realized only if genomic regions that are responsible for various traits of medical importance, such as response to a drug or susceptibility to a disease, can be identified. After the initial sequencing of the human genome, therefore, a second phase of human genomics emerged, focusing on identifying genomic variations responsible for hereditary diseases and other medically relevant traits. Such genome-wide association studies (GWAS) are based on examining the genomes of thousands of individuals for correlations between the presence of genomic variants and the trait of interest.

Many successes have come out of GWAS[5,6], but there has also been some disappointment that perhaps the pickings from these studies have been too slim[7]. For instance, although certain disorders — including obesity, diabetes and cardiovascular disease — are known to have a strong genetic component, their associated genomic variants detected through GWAS cannot explain most of the experimentally identified genetic effects found in affected families. Human geneticists call this problem the 'missing heritability'[7].

There are many possible explanations for the missing heritability, the most popular being the effect of rare variants. GWAS are based on examining a battery of different variants across the genome. Until recently, however, the cost of including both common and rare variants in such studies was prohibitively high, pushing the focus towards identifying common variants that occur at a relatively high frequency in the population. Consequently, if many rare variants, rather than a few common ones, are responsible for a disease, the rare variants would have been missed in most GWAS.
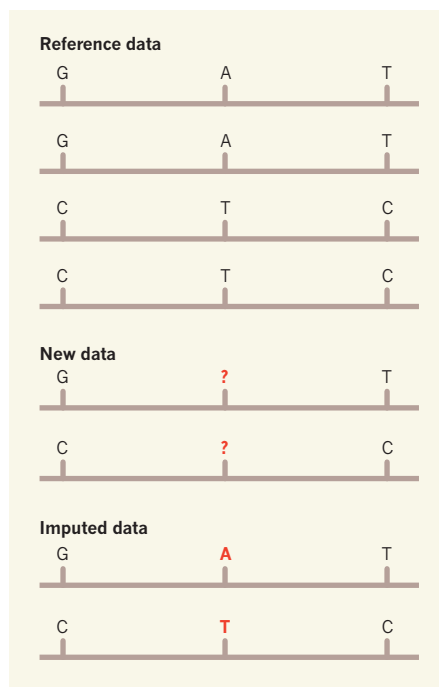
Figure 1 | Gene sequencing by imputation. On the basis of the pattern in a set of reference sequences, the missing nucleotides (indicated by question marks) in a new data set can be imputed. For example, because all sequences in the reference data with a G and a T in the first and third positions, respectively, have an A in the second position, the missing nucleotide in the first sequence of the new data is likely to be an A. Imputation methods are an integral component of the paper[2] reporting the pilot phase of the 1000 Genomes Project.

An obvious solution to this problem is to sequence whole genomes. But this is easier said than done: GWAS require sample sizes of thousands, making whole-genome sequencing extremely expensive. However, computational-biology studies have provided crucial insight that is helping to pave the way for more-comprehensive genomic studies. The idea is that if most of both common and rare variants can be characterized in just a few individuals through whole-genome sequencing, a relatively small battery of variants could then be identified in the remaining individuals in the genome-wide association study, and the pattern of those variants could be inferred computationally on the

basis of the few whole-genome sequences.

Sceptics may find this notion — using the data from some individuals to 'invent' data for others — alarming. But if done correctly, this method, called imputation[8], can significantly increase the statistical power of GWAS (Fig. 1). This idea is one of the main motivating forces behind the 1000 Genomes Project.

In the pilot phase of the project[2], the authors used several techniques to sequence the whole genomes of 179 individuals. They thereby generated a catalogue of 8 million previously unknown variants affecting single nucleotides — the building blocks of genes — and around 1 million structural variants due to small insertions or deletions of DNA. The study also presents several new methods for analysing genomic data. For example, it convincingly shows that imputation methods can significantly increase the power of GWAS.

New technologies also allow the protein-coding sequences (exons) within genes to be sequenced specifically. The vast majority of genomic DNA falls outside genes, but many of the most important variants are thought to be located within exons. Exon sequencing therefore provides a cost-effective method for identifying most of the functional variants. The consortium[2] reports exon sequences of 697 individuals from different ethnic groups.

Apart from exon sequencing, another way to contain the cost of sequencing based on GWAS is to sequence genomes at only low coverage. This means that, for each individual, only a limited amount of randomly distributed DNA is sequenced. Although, on average, a genome is sequenced several times using this technique, there may be missing data in any particular genomic region. In fact, low coverage was the approach taken for whole-genome sequencing of the 179 individuals[2].

A disadvantage of low-coverage sequencing is a higher error rate; but this can be reduced, again using imputation methods. Indeed, the consortium's low-coverage data produced an overall error rate of only 1–3% thanks to supplementation with such methods. Imputation-based methods may therefore also be the key to maximizing the utility of low-coverage sequencing data. Characterizing variants in heterozygous sites, which contain two versions

of the DNA, is more difficult, and for them the error rate in the present study varied between 5% and 30% depending on the frequency of the variant.

Given the declining cost of DNA sequencing, future discoveries in human genomics are more likely to be based on a combination of exon sequencing and low-coverage, whole-genome sequencing, rather than on the more traditional techniques. Such DNA sequencing gives access to rare and novel variants, as well as being more suitable for identifying DNA insertions and deletions and, in general, for detecting less-common variants that affect only a single nucleotide.

The remaining question is how to accommodate errors in low-coverage sequencing, because an error rate of even a few per cent can lead to drastically reduced power if not accounted for appropriately[9]. Statistical methods that incorporate high error rates will be an essential component of future

genomic-sequencing efforts. But no matter which protocol is used, the focus of the third phase of human genomics will clearly be on whole-genome sequencing. ∎

Rasmus Nielsen *is in the Departments of Integrative Biology and of Statistics, University of California, Berkeley, Berkeley, California 94720, USA.*
e-mail: rasmus_nielsen@berkeley.edu

1. www.1000genomes.org/page.php?page=about
2. The 1000 Genomes Project Consortium *Nature* **467,** 1061–1073 (2010).
3. International Human Genome Sequencing Consortium *Nature* **409,** 860–921 (2001).
4. Venter, J. C. *et al. Science* **291,** 1304–1351 (2001).
5. Hindorff, L. A. *et al. Proc. Natl Acad. Sci. USA* **106,** 9362–9367 (2009).
6. Hardy, J. & Singleton, A. *N. Engl. J. Med.* **360,** 1759–1768 (2009).
7. Manolio, T. A. *et al. Nature* **461,** 747–753 (2009).
8. Servin, B. & Stephens, M. *PLoS Genet.* **3,** e114 (2007).
9. Huang, L. *et al. Am. J. Hum. Genet.* **85,** 692–698 (2009).

DRUG DEVELOPMENT

Longer–lived proteins

Short residence times in the bloodstream reduce the effectiveness of protein drugs. Application of an approach that combines protein and polymer engineering prolongs circulation time and increases drug uptake by tumours.

JEFFREY A. HUBBELL

The past 25 years have seen an explosion in the number of approved protein drugs produced by genetic engineering, for treating hormonal, metabolic, immunological, haematological and reproductive disorders, as well as cancer[1]. Scientists initially sought to perfectly copy nature's structural expression of these proteins, leading to many first-generation drugs. Subsequently, protein engineers began to adapt nature's structures, either subtly (for example, by changing a few amino-acid residues to make interactions with a target molecule stronger or more specific) or more profoundly (for instance, by attaching two unrelated proteins to create a protein possessing a combined function that nature never considered). Several second-generation drugs have resulted from such efforts.

One drawback of protein drugs is their rapid clearance from the systemic circulation. Writing in *Proceedings of the National Academy of Sciences,* Chilkoti and colleagues[2] now describe a combined protein- and polymer-engineering approach to prolong protein circulation and enhance drug accumulation in tumours.

The concept of polymer attachment to proteins first arose in the late 1970s, with the demonstration[3] that conjugation of multiple copies of a relatively low-molecular-weight, water-

soluble, nonionic polymer, poly(ethylene glycol) (PEG), could prolong the circulation of a therapeutic enzyme. This observation led to a flurry of activity in the 'PEGylation' of protein drugs, several of which have now entered the marketplace[4–6].

An example that illustrates both the benefits and the complexities of PEGylation is interferon-α2a. The drug has been grafted at amine groups on lysine amino-acid residues to a branched, 40-kilodalton PEG chain. Although the protein is grafted with only one polymer chain, the chain can be attached to any one of four sites; as such, the drug is a mixture of four isomers. Grafting increases the hydrodynamic radius, making the drug bulkier to promote longer retention in the circulation[7]. PEGylated interferon-α2a is a very successful drug for treating chronic hepatitis C.

Polymer grafting to protein drugs is associated with many complexities, however, which Chilkoti and colleagues target in their work[2]. One such complexity, as mentioned above, is the possibility of multiple sites of polymer conjugation, leading to a heterogeneous product. A second results from limitations in the size of the polymer chain that can be grafted. Just as it is difficult to find the end of a long rope piled up in a heap, it is difficult to graft the terminus of a long polymer to the surface of a protein. An alternative approach, which is being
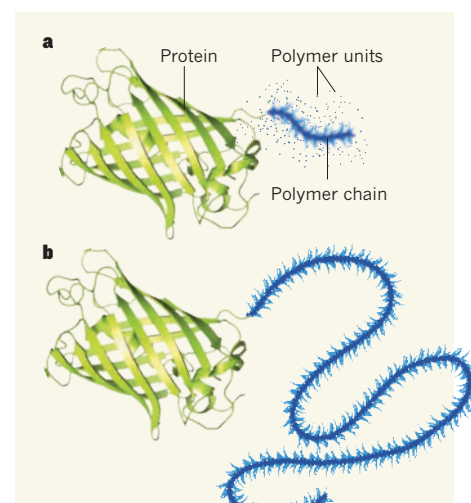
Figure 1 | Drug modification with a feather boa. Chilkoti and colleagues' approach[2] involved growing a long polymer chain from the carboxy terminus of a model protein (green fluorescent protein). The polymer unit used by the authors was oligo(ethylene glycol) methyl ether methacrylate. It is difficult to attach a long polymer chain to the end of a protein, because the two reactive sites only rarely find each other. But a chain much larger than the protein itself can be readily grown by polymerization.

developed by Chilkoti and colleagues[2,8], is to grow the polymer chain on the protein drug by polymerization; this would, in principle, allow any length of polymer chain to be grafted.

Chilkoti and colleagues' strategy[2] has many advantages. To solve the problem of multiple sites of polymer grafting, they used a protein-engineering trick to place a single chemical group at the carboxy terminus of the protein. They used this group to attach an initiator molecule for a polymerization reaction, selecting a strategically advantageous initiator for a polymerization reaction that gives precise control of polymer length under mild conditions, consistent with the delicate nature of proteins. This allowed the growth of a very long polymer chain, one that looks like a bottlebrush, consisting of a long main chain covered by short PEG chains along its length, with the polymer attached to the protein's carboxy-terminal site.

The result of this convergence of protein and polymer engineering was anything but subtle: the bottlebrush polymer on the carboxy terminus of the model protein increased its hydrodynamic radius almost sevenfold, from 3 to 20 nanometres — an increase in size corresponding to an almost 300-fold increase in hydrodynamic volume. One can imagine the result as being rather like an elfin dancer adorned with an outrageously long and fluffy feather boa (Fig. 1), rather than a few peacock feathers, as would be the effect using previous approaches. The grafted polymer chain resulted in a substantial prolongation in circulation time, which the authors showed to be beneficial in targeting tumours.

The microvasculature of tumours is known to be leaky compared with that of most normal tissues, and this difference has been used for targeting polymer-conjugated drugs — they leak slowly from healthy microvessels, but quickly from the microvessels supplying tumours[5]. However, if the drug does not circulate for long, it has little time to leak from the tumour microvessels. The 20-nm polymer–protein conjugate did indeed experience longer circulation times, and it leaked into tumours in mice 50 times more efficiently than the unmodified protein[2]. For human treatments, it would be necessary to engineer an actual drug with the polymer graft, rather than a model protein as used here, and to show that the terminally attached polymer does not disrupt drug function.

Therapeutic proteins have already had a tremendous impact on human health, and protein–PEG conjugates make up a substantial minority of them. But the simultaneous engineering of the protein for the polymer and the polymer for the protein — as implemented by Chilkoti's team — will open further doors for developing protein drugs. ∎

Jeffrey A. Hubbell *is at the Institute of Bioengineering and the Institute of Chemical Sciences and Engineering, Ecole Polytechnique Fédérale de Lausanne (EPFL), CH-1015 Lausanne, Switzerland.*
e-mail: jeffrey.hubbell@epfl.ch

1. Leader, B., Baca, Q. J. & Golan, D. E. *Nature Rev. Drug Discov.* **7**, 21–39 (2008).
2. Gao, W., Liu, W., Christensen, T., Zalutsky, M. R. & Chilkoti, A. *Proc. Natl Acad. Sci. USA* **107**, 16432–16437 (2010).
3. Abuchowski, A., McCoy, J. R., Palczuk, N. C., van Es, T. & Davis, F. F. *J. Biol. Chem.* **252**, 3582–3586 (1977).
4. Duncan, R. *Nature Rev. Drug Discov.* **2**, 347–360 (2003).
5. Duncan, R. *Nature Rev. Cancer* **6**, 688–701 (2006).
6. Veronese, F. M. & Pasut, G. *Drug Discov. Today* **10**, 1451–1458 (2005).
7. Bailon, P. *et al. Bioconj. Chem.* **12**, 195–202 (2001).
8. Gao, W. *et al. Proc. Natl Acad. Sci. USA* **106**, 15231–15236 (2009).

BIOGEOCHEMISTRY
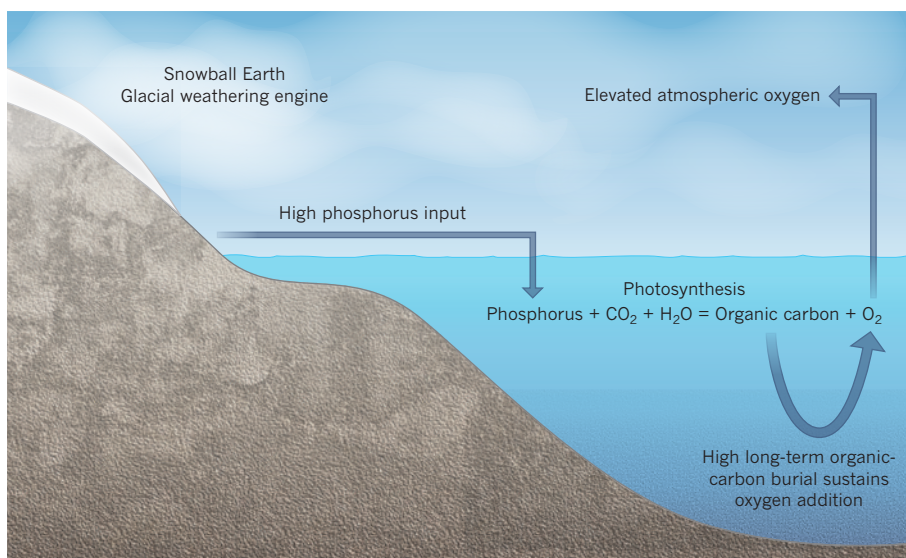
Phosphorus and the gust of fresh air

Evidence of intense phosphorus weathering following 'snowball Earth' glaciations raises a further possibility — that this revved-up nutrient cycle drove conditions for the explosion of animal life. SEE LETTER p.1088

GABRIEL M. FILIPPELLI

The rapid diversification of animal (metazoan) life that started at the end of the late Proterozoic eon, about 700 million years ago, marked a turning point in Earth's biological systems. The increase in atmospheric oxygen that occurred around this time certainly had something to do with this, by providing adequate oxidant for respiration and a sufficient stratospheric ozone layer for protection from ultraviolet radiation[1,2]. But perhaps the end of the widespread 'snowball Earth' glaciations, which covered most of the land surface and oceans in ice, may have been a factor[3]. Could the rapid increase in atmospheric oxygen be somehow related to these glaciations?

In this issue (page 1088), Planavsky *et al.*[4] exploit analyses of the ratio of phosphorus to iron in ancient marine iron-oxide deposits to point an accusatory finger at phosphorus as the connecting agent. This element dictates global biological productivity, and consequently the burial of organic carbon in the ocean on geological timescales[5]. Past attempts to constrain the history of the phosphorus cycle have been hampered by a lack of comprehensive understanding of phosphorus weathering, transport, recycling and ultimate burial in marine sediments. Recent work[6–8], however, has bolstered confidence that the sedimentary record of phosphorus can be used to interpret changes in biological productivity and global phosphorus cycling on geological timescales.

Planavsky *et al.*[4] have now used the phosphorus/iron ratio in iron-oxide-rich sedimentary rocks to constrain estimates of the dissolved phosphorus concentration in ancient sea water. In the laboratory, iron oxides and phosphorus co-precipitate with a characteristic phosphorus/iron ratio that is primarily related to the phosphorus concentration of the precipitating fluid. The process is also influenced by competing elements such as silicon, however, and translating beaker-scale experiments into deep time is problematic, requiring careful selection of candidate rocks and analysis of various post-depositional factors that could destroy the integrity of the phosphorus/iron relationship. The authors mined the literature and performed analyses of their own, and then removed the silicon signal by making assumptions about past marine silicate concentrations drawn from palaeo-ecological reconstructions.

What Planavsky *et al.* found was remarkable — relatively 'normal' and constant phosphorus concentrations from 3 billion to 1.5 billion years ago, then (following a significant data gap) a huge peak reaching ten times

Figure 1 | Snowball Earth glaciations and the rise in oxygen levels around 700 million years ago[4]. Increased chemical weathering of terrestrial phosphorus resulted from the physical weathering caused by glacial processes. This revved-up phosphorus cycle was sustained by the lack of soil formation owing to the absence of plant rootedness, and meant that phosphorus bled off the land in bioavailable forms instead of being trapped in organically and oxide-bound forms. The continual high input of phosphorus, a biolimiting nutrient, drove intensified production of organic matter. The ensuing enhanced burial of organic carbon in turn drove addition of oxygen to the ocean and the atmosphere through their mutual long-term mass balances[9] — so, speculatively, providing conditions that were ultimately conducive to metazoan evolution.

its current level around 700 million years ago, coinciding with the cessation of snowball Earth conditions. This peak reflects a large increase in the marine phosphorus inventory that was sustained over tens of millions of years. Given that 'modern' marine phosphorus has a residence time of tens of thousands of years, it is hard to imagine ocean conditions in which sustained high levels of phosphorus were not driven by a step-change in the global phosphorus mass balance.

The thinking about the consequences of such high phosphorus levels then runs as follows. The result of the phosphorus-driven marine productivity was sustained algal blooms in the ocean, much like those found today in ponds and streams near areas of heavy fertilizer application. The death and settling of these blooms caused long-term, enhanced organic-carbon burial, which (via the mass-balance relationship between carbon and oxygen[9]) resulted in the addition of oxygen to the ocean–atmosphere system (Fig. 1). This increase in atmospheric oxygen controlled the evolutionary patterns of oxygen-dependent metazoans. Such a scenario provides a plausible link between the roles of snowball Earth glaciations and late Proterozoic oxygenation in leading to the explosion in metazoan diversity.

The implications of the new results[4] for understanding glacially induced phosphorus weathering on landscapes are also interesting. We are beginning to appreciate how glacial dynamics affects phosphorus weathering on land and transport to the oceans[10]. In the modern 'rooted' world, in which soil development is generally mediated by plants, most of the weathered phosphorus is mobilized and transported from landscapes in a narrow time window after a glacier retreats. Continental records[11] indicate that this large flux of phosphorus occurs in about 10,000 years in most landscapes (probably faster in low-relief/high-rainfall landscapes and slower in high-relief/low-rainfall landscapes). In a modern landscape with its considerable plant coverage, maintenance of a sustained increase in phosphorus delivery to the oceans would require periodic removal of weathered and phosphorus-depleted soils, and exposure of fresh material for further soil development.

In late Proterozoic time, however, the absence of land plants meant that there would have been little soil development to stabilize landscapes, or to convert mineral-based phosphorus forms on pristine mineral surfaces to the organically and oxide-bound phosphorus found in modern soils. Presumably, phosphorus stripping from rocks was much more extensive, as the landscapes themselves were much less stabilized because of the lack of flora. Thus, in this case, the present is not a key to the distant past. Without the systems to stabilize phosphorus, the phosphorus cycle in the late Proterozoic would have been

permanently revved up — that is, until rootedness came into play several hundred million years later. A test of this hypothesized mechanism of enhanced phosphorus stripping from landscapes would be to identify deltaic or other sedimentary-basin environments from the late Proterozoic, and to use proxy estimates of phosphorus loss (for example, the phosphorus/aluminium ratio) to determine whether values for this time interval are lower than expected given the source material. Such analyses would provide independent evidence of high phosphorus weathering rates.

Meanwhile, with this paper[4] and use of the phosphorus/iron proxy, there is now another way to look at nutrient variations in the ancient oceans. No proxy is perfect, however. In this instance, the weaknesses are both obvious (the limited spatial and temporal range of the proper rock types for analysis) and less obvious (variations in original iron-oxide composition that are now masked by mineral maturation). Nevertheless, thanks to Planavsky and colleagues[4], we have a picture of the marine phosphorus cycle through deep time. We can

begin to develop informed hypotheses about how variations in the phosphorus cycle are driven, and what impact they have on the global carbon cycle, oxygen levels and the evolution of marine ecosystems. ∎

Gabriel M. Filippelli *is in the Department of Earth Sciences, Indiana University–Purdue University Indianapolis (IUPUI), Indianapolis, Indiana 46202, USA.*
e-mail: gfilippe@iupui.edu

1. Catling, D. C., Glein, C. R., Zahnle, K. J. & McKay, C. P. *Astrobiology* **5,** 415–438 (2005).
2. Hoffman, P. F., Kaufmann, A. J., Halverson, G. P. & Schrag, D. P. *Science* **281,** 1342–1346 (1998).
3. Hoffman, P. F. & Schrag, D. P. *Terra Nova* **14,** 129–155 (2002).
4. Planavsky, N. J. *et al. Nature* **467,** 1088–1090 (2010).
5. Tyrell, T. *Nature* **400,** 525–531 (1999).
6. Mort, H. P. *et al. Geology* **35,** 483–486 (2007).
7. Paytan, A. & McLaughlin, K. *Chem. Rev.* **107,** 563–576 (2007).
8. Diaz, J. *et al. Science* **320,** 652–655 (2008).
9. Berner, R. A. *Nature* **426,** 323–326 (2003).
10. Föllmi, K. B., Hosein, R., Arn, K. & Steinmann, P. *Geochim. Cosmochim. Acta* **73,** 2252–2282 (2009).
11. Filippelli, G. M. *et al. Quat. Res.* **66,** 158–166 (2006).

CANCER

Genomic evolution of metastasis

Prognosis for patients with pancreatic cancer is bleak, often owing to late diagnosis. The estimate that at least 15 years pass from tumour initiation to malignancy offers hope for early detection and prevention. SEE LETTERS P.1109 & P.1114

E. GEORG LUEBECK

Radiocarbon dating and comparative analyses of skeletal anatomy have informed the theory of human evolution; similarly, DNA sequencing of tumour cells coupled with dissection of the molecular anatomy of chromosomal aberrations is beginning to yield deeper insight into the evolution of cancer. In this issue, two papers[1,2] present findings from sequencing the protein-coding regions (exons) of more than 20,000 genes from the genomes of patients with metastatic (stage IV) pancreatic cancer. The findings are unprecedented, providing the first high-resolution image (at the level of single base pairs) of the non-germline mutational spectrum of pancreatic tumours and their metastatic descendants.

It has long been recognized that genomic instability is a hallmark of cancer. However, its significance in cancer progression has been the subject of debate for just as long. To shed light on this, Campbell *et al.*[1] (page 1109) performed a DNA-sequence-based study of chromosomal

rearrangements. They find that specific chromosomal rearrangements known as fold-back inversions occur in almost all of a patient's metastatic lesions. What's more, unlike other chromosomal rearrangements, which seem to occur either in the primary, parental tumour or in the metastatic lesion, Campbell *et al.* detect fold-back inversions in both primary and metastatic tumours. The authors therefore argue that fold-back inversions occur early in tumorigenesis and are probably a crucial driver of pancreatic-cancer progression.

The precise origin of fold-back inversions is unknown. It could be that DNA-replication-related erosion of the telomeres (the chromosome ends) — potentially because of suppressed or dysfunctional activity of the enzyme telomerase — triggers recurrent breakage–fusion–bridge (B/F/B) cycles[3,4], which, in turn, cause progressive gains and losses of genetic material and so genomic instability. Intriguingly, telomerase activity seems to be restored in the invasive tumours, which might have a stabilizing effect on the abundance of B/F/B-induced rearrangements,
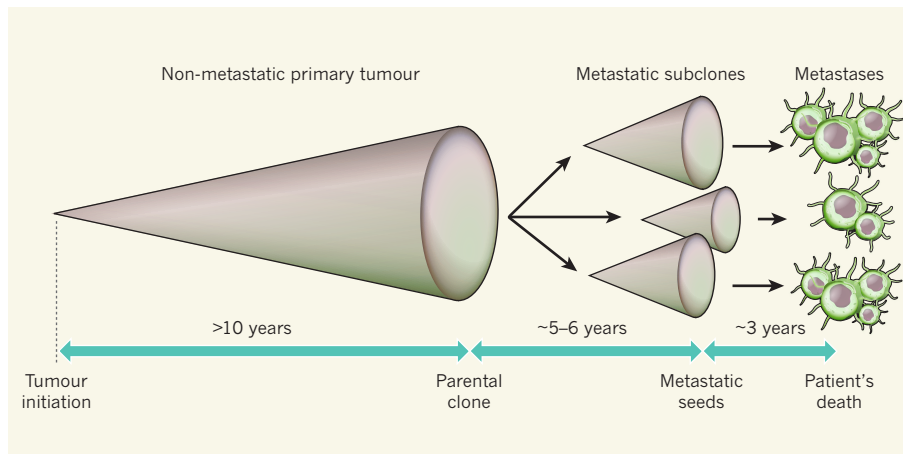
Figure 1 | The pancreatic-cancer timeline. Mathematical analyses of tumour-DNA sequence data from two collaborative studies[1,2] suggest that it probably takes more than 10 years from the initiation of a pancreatic tumour to the birth of the parental clone that results in pancreatic cancer. However, this clone does not have metastatic potential, and the subclones with the ability to spread to other tissues develop over an additional 5–6 years. The metastases, which are soon followed by the patient's death, occur over roughly the next 3 years.

but not on other rearrangements.

Campbell and colleagues' results affirm the presence of genomic instability in the development of pancreatic cancer. But because of extensive differences in the number, type and position of the rearrangements among patients — and even between the metastatic deposits in the same organ of a single patient — the functional consequences of this instability remain unclear. Studies using next-generation sequencing technologies on a larger number of patients are likely to fill in the missing pieces and pinpoint the driving forces in tumour progression and metastatic dissemination across different types of cancer.

In a separate study, Yachida *et al.*[2] (page 1114) address the clinically relevant issue of the timescales associated with tumour progression. These authors also carry out genomic sequencing of pancreatic-cancer metastases and examine their phylogenetic relationship with their respective, previously sequenced, primary tumours in seven patients. They thus derive estimates of three timescales: the time from tumour initiation to the birth of the founder cell of the parental (non-metastatic) clone; the sojourn time between the parental clone arising and its acquisition of metastatic potential; and the time from metastatic dissemination to the patient's death (Fig. 1).

Remarkably, the authors estimate that the time from tumour initiation to metastatic dissemination is at least a decade — a conclusion that suggests that there is a window of opportunity for medical intervention before the cancer spreads to distant organs. This finding is not inconsistent with that inferred from quantitative analyses of the age-specific incidence of pancreatic cancer in the general population[5]. On the basis of a general mathematical description that recognizes the random nature of both mutation accumulation and clonal expansion in pancreatic cancer, the

earlier, population-based analysis[5] estimated that the mean sojourn time from the tumour-initiating mutation to clinical diagnosis may be as much as five to six decades.

On the surface, the population-based estimate seems much longer than Yachida *et al.* conclude. It should be kept in mind, however, that the present sequence-based time estimates[2] are not general: they do not refer to the average of all pancreatic lesions with cancerous and metastatic potential in the tissue, but rather refer to the one lesion in the tissue which, by chance, leads to the first primary tumour in

that tissue. Thus, Yachida and colleagues' estimate must be considered a lower bound for the mean sojourn time of pancreatic lesions, such as pancreatic intraepithelial neoplasia, that have the potential to cause invasive and metastatic cancer. From a clinical perspective, what matters is the prospective disease risk, which may involve multiple lesions individually evolving towards cancer. Thus, the time estimates of Yachida *et al.* are conservative and so clinically relevant.

These two studies[1,2] are a bellwether, and are among the first to explore the biological and clinical implications of sequence data for individual tumours. As the sequencing technology moves forward — and it does so at a blinding speed — more exciting details of the evolutionary processes involved in tumour progression are likely to be unearthed. It is to be hoped that such information will not only deepen our understanding of the cancer process, but also lead to new approaches to early cancer detection, better prognosis and, ultimately, prevention. ∎

E. Georg Luebeck *is at the Fred Hutchinson Cancer Research Center, Program in Computational Biology, Seattle, Washington 98109-1024, USA.*
e-mail: gluebeck@fhcrc.org

1. Campbell, P. J. *et al. Nature* **467,** 1109–1113 (2010).
2. Yachida, S. *et al. Nature* **467,** 1114–1117 (2010).
3. McClintock, B. *Genetics* **26,** 234–282 (1941).
4. Lo, A. W. I. *et al. Neoplasia* **4,** 531–538 (2002).
5. Meza, R., Jeon, J., Moolgavkar, S. H. & Luebeck, E. G. *Proc. Natl Acad. Sci. USA* **105,** 16284–16289 (2008).
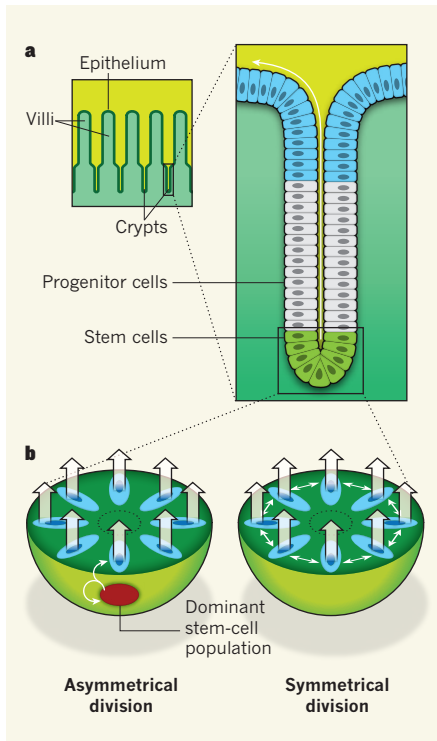
STEM CELLS

The intestinal–crypt casino

Stem cells can renew themselves indefinitely — a feature that is often attributed to asymmetrical cell division. Fresh experimental and mathematical models of the intestine provide evidence that begs to differ.

MICHAEL P. VERZI & RAMESH A. SHIVDASANI

Certain tissues, such as the skin, blood and intestinal lining, replenish millions of lost cells every day. The burden of renewal falls on small populations of stem cells, which can make exact copies of themselves, as well as generate all the resident cell types that differentiate and eventually die. This rare dual ability, a defining property of all stem cells, is exemplified by the asymmetrical division of germ cells[1] and neuronal precursors[2] in the fruitfly. Nonetheless, as long as the total stem-cell pool in a tissue remains roughly constant,

in principle there is no reason why individual stem cells should not divide symmetrically, to generate either two identical stem cells or two daughters that exit the pool to differentiate. Indeed, two reports published in *Science*[3] and *Cell*[4] demonstrate that, in the normal course of tissue renewal, intestinal stem cells divide symmetrically.

In the small intestine, stem cells lie at visually identifiable positions within pocket-like crypts, and their progeny migrate in predictable streams (Fig. 1a, overleaf). In mice, two cell populations manifest the capacity for both prolonged self-renewal and multi-lineage

Figure 1 | Intestinal stem cells and their renewal. a, In the small intestine, a sheet of epithelial cells lines innumerable finger-like projections called villi. Short invaginations of the epithelium between villi, called crypts, house intestinal stem cells, which replace the lost epithelial cells of the villi. **b**, There are two alternative models of how intestinal stem cells replicate to replenish the lost cells. The asymmetrical-division model proposes that a dominant stem cell (red) hierarchically gives rise to other cells (blue) through asymmetrical cell division, producing a copy of itself every time it divides. However, two new studies[3,4] propose that, in the intestinal crypt, stem cells that are roughly equal in most respects can divide symmetrically to replace their neighbours. Thin arrows depict cell divisions that give rise to stem cells; thick arrows depict divisions that produce differentiated cells. (**b** adapted from ref. 3.)

differentiation: rare cells in the fourth tier of the crypt that express the nuclear protein Bmi1 and replicate infrequently[5], and more abundant cells — about a dozen — that lie deep in the crypt base, express the surface protein Lgr5 and divide almost daily[6]. Whereas both of these cell populations seem to replenish the intestinal lining repeatedly and indefinitely, only the Lgr5-expressing cells have been shown[7] to generate limitless stem-cell progeny in the laboratory.

It is possible to mark individual cells in an animal with an enzyme or fluorescent protein and then to follow the label to monitor those cells' descendants. Such an approach previously established[8] that intestinal crypts are monoclonal — that is, all their resident cells derive from the same parent cell. This finding indicated that the progeny of certain stem cells

can exclusively populate a crypt, replacing even other stem cells within that crypt.

Lopez-Garcia et al.[3] and Snippert et al.[4] exploit this clonal property to ascertain fundamental features of stem-cell replication. Shortly after genetic labelling in adult mice, a few crypt cells carried the tracer, signalling their recent origin in a marked progenitor. Over time, the fraction of marked cells in individual crypts increased or decreased until all crypt cells were uniformly labelled or unlabelled, reflecting their derivation from a stem cell, or stem cells, that had replaced all others. Integral to both studies was a means of tracking cells quantitatively during their 3–4-month journey towards becoming monoclonal.

The investigators' models allowed them to ask a crucial question: are all stem cells within a crypt equally poised to divide symmetrically and randomly, or does a dominant stem cell rule a hierarchy — by dividing asymmetrically to copy itself and simultaneously produce a secondary stem cell of lower rank, one that can give rise only to cells that differentiate (Fig. 1b)? They hypothesized that if the genetic label landed by chance in a dominant stem cell, over time the mark would spread to every other stem cell in that crypt. Chance labelling of secondary stem cells, however, would result in their eventual displacement by the dominant cell's unmarked or differently marked progeny. Conversely, if many crypt stem cells have roughly similar potential to divide symmetrically, clones originating in single labelled cells should grow and shrink in size until they either dominate the crypt or become extinct through equal competition from cells with a different label.

This scenario resembles one in which gamblers with equal, small purses place successive wagers, each with equal odds. Although no gambler has an intrinsic edge, the one fortunate enough to win a few early rounds by chance enlarges her purse enough to raise the subsequent ante, putting her in a favourable position to overtake the competition. Like many gambling games, this one, with its counterpart in intestinal crypt clonality, can be modelled mathematically[9], allowing the pattern of clone-size evolution to distinguish between alternative mechanisms of stem-cell replication (Fig. 1b).

Random division of stem cells with roughly equal potential should give clone-size distributions that converge on a pattern known as scaling — a cardinal feature of random games[9]. If, however, stem cells derive from a single dominant source, labelling of that source should yield clones that grow steadily, whereas labelling of lower-ranking cells should yield clones that eventually disappear because the progeny of unlabelled (or differently labelled) dominant cells takes over; clone sizes in this case would show a binomial distribution rather than scaling properties.

Data from different labelling methods in

the two laboratories[3,4] uncovered unambiguous scaling behaviour in clone sizes. Although the results cannot exclude the occasional surge of hierarchy, they indicate that, under normal conditions over several months, intestinal stem cells compete continuously on the level field of symmetrical replication. Tangentially, they also imply that Bmi1-expressing and Lgr5-expressing cells may represent a largely overlapping pool of stem cells, as investigators have begun to suspect[10].

An excess of stem cells could cause tissue overgrowth or cancer, whereas a deficit may contribute to ageing and organ failure. Organ renewal therefore depends on striking the right balance between retention and surrender of stem cells' replicative ability. Asymmetrical division allows intrinsically strict control of stem-cell numbers and also limits the amount of DNA damage that continually renewing cells might propagate. These are some of the reasons for the long-standing appeal of asymmetrical division as an inherent stem-cell property. Brisk, symmetrical turnover of gut stem cells potentially forfeits these advantages, suggesting the presence of other safeguards against indefinite expansion of compromised cells.

Future experiments might address the mechanisms that determine and maintain stem-cell numbers in intestinal crypts. In other words, how does each intestinal stem cell first perceive the need to divide and then choose between spawning two similar new stem cells or two daughters that permanently exit the stem-cell pool through differentiation? The answer to this question could help in devising fundamentally new cancer treatments and in harnessing the promise of regenerative medicine. The odds favour intestinal crypts as a source of incisive solutions. ■

Michael P. Verzi and **Ramesh A. Shivdasani** are in the Department of Medical Oncology, Dana-Farber Cancer Institute, Boston, Massachusetts 02115, USA. They are also in the Department of Medicine, Brigham & Women's Hospital and Harvard Medical School.
e-mail: ramesh_shivdasani@dfci.harvard.edu

1. Fuller, M. T. & Spradling, A. C. *Science* **316**, 402–404 (2007).
2. Knoblich, J. A. *Cell* **132**, 583–597 (2008).
3. Lopez-Garcia, C., Klein, A. M., Simons, B. D. & Winton, D. J. *Science* doi:10.1126/science.1196236 (2010).
4. Snippert, H. J. *et al. Cell* **143**, 134–144 (2010).
5. Sangiorgi, E. & Capecchi, M. R. *Nature Genet.* **40**, 915–920 (2008).
6. Barker, N. *et al. Nature* **449**, 1003–1007 (2007).
7. Sato, T. *et al. Nature* **459**, 262–265 (2009).
8. Winton, D. J., Blount, M. A. & Ponder, B. A. J. *Nature* **333**, 463–466 (1988).
9. Bramson, M. & Griffeath, D. *Probab. Theor. Relat. Fields* **53**, 183–196 (1980).
10. Li, L. & Clevers, H. *Science* **327**, 542–545 (2010).

ASTROPHYSICS

Weighing in on neutron stars

The more massive a neutron star is, the greater the constraints it places on the nature of the matter at its core. The discovery of a new mass record holder has strengthened those constraints considerably. SEE LETTER P.1081

M. COLEMAN MILLER

Large mass is a touchy subject among humans, but for neutron stars it is greatly desirable. This is because high mass places strong constraints on the matter in these stars' cores, which exists in a state that cannot be probed in laboratories and could be dominated by anything from neutrons and protons to exotica (such as quark matter that is not confined inside nuclei, hyperons or condensates). On page 1081 of this issue, Demorest et al.[1] report measurements of a neutron star with a mass nearly 20% greater than any previous, precisely measured value.

The object studied by Demorest and colleagues is a millisecond pulsar with a companion star. A millisecond pulsar is a rotating neutron star that emits a beam of radio waves at regular millisecond intervals. For this particular source, known from its sky coordinates as J1614−2230, the authors use a felicitous orientation of the binary system, along with the extreme timing stability characteristic of millisecond pulsars, to measure the 'Shapiro delay'. This delay, predicted in 1964 by Irwin Shapiro[2] using general relativity, occurs when

light passes through the gravitational field of an object during its journey to Earth (Fig. 1).

As seen from a large distance, clocks run more slowly in deeper gravitational potentials. As a result, in a binary system this delay increases and decreases periodically with a characteristic shape and depth that depend, respectively, on the system's orbital inclination relative to our line of sight and the mass of the companion. When combined with classical measurements of the binary orbital period and of the line-of-sight speed of motion of the pulsar (as determined from Doppler timing shifts), the delay yields the masses of both the pulsar and its companion. The Shapiro delay has been measured before (for example, for the double pulsar J0737−3039; ref. 3), but never with enough accuracy to provide precise mass estimates without measurements of additional relativistic parameters.

Importantly, and unlike alternative post-Keplerian effects such as the precession of the orbital pericentre (the point at which the two masses are closest), the Shapiro delay does not depend on complicating effects such as tidal forces on the companion. It therefore provides a robust estimate of the masses. It
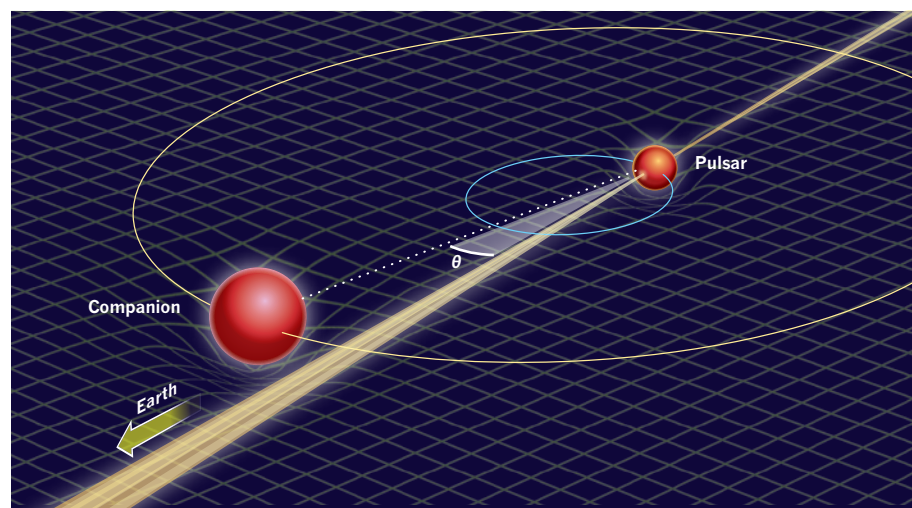
Figure 1 | The Shapiro delay. Radio waves from a pulsar that pass close to a companion are affected by time dilation through the companion's gravitational well. At a given moment, the resulting time delay due to a companion of gravitational mass M is $\Delta t = -(2GM/c^3)\ln(1-\sin i \cos\theta)$, where G is Newton's constant, c is the speed of light, i is the inclination of the orbit to our line of sight, and θ is the instantaneous phase of the orbit. Nearly edge-on systems such as the binary pulsar J1614−2230 studied by Demorest et al.[1] produce comparatively large Shapiro delays during conjunction (where $\theta = 0$).

is, however, a weak effect, which is difficult to measure unless the binary system is nearly edge-on to us. In such a configuration, the radio waves from the pulsar pass very close to the companion at conjunction (the point at which the objects are nearest to each other on the sky), thus maximizing the signal. Indeed, J1614–2230 is the most edge-on binary millisecond pulsar known, with an orbital inclination of about 89.17°.

Demorest *et al.*[1] find that the neutron star in J1614–2230 has a gravitational mass (that is, the mass that would be measured by a distant satellite in orbit, in contrast to the approximately 20–30% larger 'baryonic mass' that would be obtained by summing the masses of all the object's constituent particles) of 1.97 ± 0.04 times the mass of the Sun. For comparison, before this, the highest precisely measured mass was 1.67 ± 0.01 times the mass of our Sun for the binary pulsar J1903+0327 (ref. 4), and the ultraprecisely measured masses of double neutron-star binaries span a range of only 1.25–1.44 times the mass of our Sun[5].

The reason that such masses matter to astrophysicists and nuclear physicists alike is that the matter in the cores of neutron stars exists in a regime that cannot be probed terrestrially, and nuclear theories on the composition and properties of this regime disagree strongly. Ordinary nuclei on Earth have densities of roughly 2.6×10^{14} times that of water. They also tend to have almost equal numbers of neutrons and protons. By contrast, neutron-star cores are thought to have densities that are about 2–10 times higher, and if they comprise primarily neutrons and protons then they have roughly ten times as many neutrons as protons.

In addition, other particles may dominate the interiors of neutron stars. The uncertainty principle of quantum mechanics says, among other things, that one cannot measure the position and momentum of a particle simultaneously and exactly. Thus, when neutrons are confined in a high-density region (and so the volume per particle is small and their positional uncertainty is low), they acquire a quantum-mechanical momentum, called the Fermi momentum, that can be substantial. The associated Fermi energy adds to the energetic cost per neutron; if other particles of lower total energy exist, they can substitute for neutrons. Hence hyperons and other exotica are a possibility in neutron-star cores, but there are enough uncertainties about, for example, their self-interactions that the ground state of high-density matter is not known.

One of the strongest discriminants between different models is the maximum observed mass among neutron stars. Models with exotica tend to have lower maximum masses than models dominated by neutrons and protons. This is because a transition to a lower-energy state at a high density makes the matter easier to compress and thus less resistant to gravity. In principle, modellers devoted to hyperons

or quark matter can adjust free parameters to account for the mass of the neutron star in J1614–2230, but almost all existing models with such compositions are ruled out for a 1.97-solar-mass star (see ref. 6 for a recent review). Demorest and colleagues' results[1] are thus a vote in favour of neutrons and protons.

Even better, we may be seeing the dawn of an era in which Shapiro-delay measurements in pulsar binaries become more common and do not require remarkable chance alignments. The current limiting factor is the measurement precision of the arrival time of the radio waves. However, this precision is being improved rapidly thanks to its anticipated application in gravitational-wave detection by pulsar-timing arrays[7], which measure tiny shifts, induced by gravitational waves, in the arrival times of radio pulses from a collection of pulsars distributed on the sky. As a result, more and

more systems will be open to such analysis, and we may see further evidence that heavy is beautiful. ∎

M. Coleman Miller *is in the Department of Astronomy, University of Maryland, and the Joint Space Science Institute, College Park, Maryland 20742–2421, USA.*
e-mail: miller@astro.umd.edu

1. Demorest, P. B., Pennucci, T., Ransom, S. M., Roberts, M. S. E. & Hessels, J. W. T. *Nature* **467,** 1081–1083 (2010).
2. Shapiro, I. I. *Phys. Rev. Lett.* **13,** 789–791 (1964).
3. Kramer, M. & Wex, N. *Class. Quant. Grav.* **26,** 073001 (2009).
4. Champion, D. J. *et al. Science* **320,** 1309–1312 (2008).
5. Thorsett, S. E. & Chakrabarty, D. *Astrophys. J.* **512,** 288–299 (1999).
6. Lattimer, J. M. & Prakash, M. *Nucl. Phys. A* **777,** 479–496 (2006).
7. Hobbs, G. *et al. Class. Quant. Grav.* **27,** 084013 (2010).

STRUCTURAL BIOLOGY

A peep through anion channels

The crystal structure of a protein channel provides clues about the mechanisms that control the closure of pores found in the epidermis of plant leaves. Excitingly, the protein channel folds in a way never seen before. SEE ARTICLE P.1074

SÉBASTIEN THOMINE
& HÉLÈNE BARBIER-BRYGOO

You might think that bacteria have little to teach us about plants. But as Chen *et al.* reveal on page 1074 of this issue[1], you would be wrong. They report the three-dimensional structure of a protein from the bacterium *Haemophilus influenzae* that is structurally similar to the ion channel SLOW ANION CHANNEL 1 (SLAC1) found in plants, which is a key regulator of gas exchange between plants and the atmosphere. The structure reveals a previously unreported protein design that allows anions to permeate through membranes. By comparing the structure of the bacterial protein with a model of SLAC1, the authors were able to make selective mutations to the plant protein to investigate its activation mechanism.

The leaf epidermis of terrestrial plants contains pores known as stomata, which are formed by two kidney-shaped guard cells (Fig. 1). The pores' role is to control gas exchange between air spaces inside the leaves and the surrounding atmosphere. The influx of carbon dioxide through stomata determines the photosynthetic efficiency of the leaves, whereas the control of water-vapour efflux

through the pores is central to maintaining water balance in plants. Drought, elevated carbon dioxide, ozone or pathogen attacks induce stomata to close[2], with the size of stomatal apertures being determined by the osmotic potential in the guard cells.

Stomatal closure is mediated by the release of ions from guard cells, a process that requires the coordinated efflux of anions and potassium ions accumulated inside those cells. The activation of anion channels is an essential step in stomatal closure, because it leads both to anion efflux and to the depolarization of cell membranes necessary to activate potassium-efflux channels. The membranes of stomatal guard cells harbour slowly activating anion channels (S-type channels), which display a strong preference for nitrate ions over chloride ions, and which allow sustained anion efflux upon activation by phosphorylation or calcium ions[2]. It has been proposed that S-type-channel activation is the key event leading to stomatal closure.

In the model plant species *Arabidopsis thaliana*, screens for mutants lacking stomatal responses to elevated carbon dioxide or ozone have identified mutations in the *SLAC1* gene[3]. *SLAC1* encodes a protein that has ten transmembrane α-helices, and whose amino-

acid sequence is related[3] to that of an anion transporter protein from the yeast *Schizosaccharomyces pombe*. Slow anion currents (the currents through S-type channels) are absent in *slac1* mutant guard cells.

Other genetic screens for *A. thaliana* mutants possessing altered stomatal responses have identified the protein kinase OPEN STOMATA 1 (OST1) as another key control of stomatal aperture size. Co-expression of SLAC1 and OST1 allows the permeation and gating properties of the slow anion currents observed in guard cells to be reconstituted in frog egg cells[4]. *SLAC1* thus encodes the ion-permeation pore of the slow anion channel, which is directly activated by OST1-mediated phosphorylation. What's more, *slac1* mutants display impaired responses to all stimuli that induce stomatal closure, confirming the central role of slow anion channels in this process.

In the absence of a crystal structure of SLAC1 itself, Chen and colleagues' crystal structure[1] of the analogous bacterial protein TehA **represents a great advance. The struc**ture reveals that TehA is trimeric, consisting of three identical subunits. Each subunit has the general structure of a ribbon wrapped around a ring (see Fig. 2 on page 1075), with every second helix placed on the inner side of the ring to form a pore. The trimer thus constitutes a 'triple barrel' channel (Fig. 1). These features of the structure define a completely new protein fold. The authors went on to construct a homology model for SLAC1, in which the amino-acid sequence of the protein was transposed onto the TehA structure. The model predicted that the plant protein has similar structural features to those of TehA.

The inner helices of TehA (and of SLAC1) include centrally located proline amino acids that generate kinks, allowing the formation of a cylindrical pore that has a quite constant diameter through the surrounding cell membrane. The pore does not contain any obvious feature that could discriminate between ions, or an anion-binding site. Accordingly, the permeability of anions through SLAC1 is governed by the hydration energies of the ions (a measure of the strength with which they bind to water molecules), as observed for other proteins that bind weakly to anions[5]. In other words, ions that bind weakly to water molecules have greater permeability through the channel. In practice, this means that SLAC1 has greater permeability to nitrate than to chloride ions.

The permeation pathway of SLAC1 is in sharp contrast to that of another ubiquitous family of anion channels and transporters, the chloride-channel (CLC) family — ion permeation in CLCs proceeds through three anion-binding sites[6]. Nevertheless, both SLAC1 and the most-studied CLC from plants, the CLCa anion/proton antiporter from *A. thaliana*, exhibit a strong preference for nitrate, which is the main anion found in plant cells[7].

The authors tested their TehA-based SLAC1
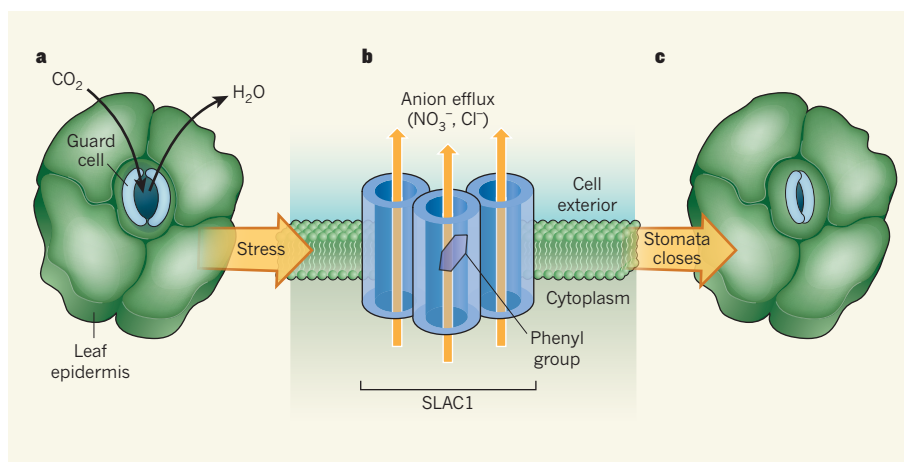
Figure 1 | Role of SLAC1 channels in stomatal closure. a, In the leaf epidermis of terrestrial plants, pores known as stomata regulate CO_2 uptake and the loss of water vapour. The pores are formed from two kidney-shaped guard cells. **b**, The SLAC1 channel in the cell membrane of guard cells is responsible for nitrate (NO_3^-) and chloride (Cl^-) efflux from the cells, a process that is triggered by stressful stimuli such as drought. Chen *et al.*[1] report the crystal structure of the bacterial protein TehA, the amino-acid sequence of which is related to SLAC1. Their model of SLAC1, based on the TehA structure, indicates that three subunits of the protein each form a pore and associate into 'triple-barrel' structures in the membrane. When closed, the pore channel of each subunit is occluded by the phenyl group of an amino-acid residue, forming the basis of a previously unknown gating mechanism common to plant and bacterial channels from the SLAC family. Here, a phenyl group is shown in an 'open' orientation; the phenyl groups in the other two subunits are not shown. **c**, The coordinated efflux of nitrate and chloride ions triggers further efflux of potassium ions and water from guard cells (not shown), resulting in stomatal closure.

model[1] by mutating either residues known to be altered in the plant *slac1* mutants[3] or key residues identified through their own structural analysis. They observed that the effects of the mutations on the conductance of the resulting SLAC1 proteins agreed with what was predicted by the model.

In both the SLAC1 model and the TehA crystal structure, the channel pore is occluded by the phenyl side chain of a phenylalanine amino-acid residue. This residue is evolutionarily conserved in all of the 900 SLAC1-related protein sequences analysed by Chen *et al.*[1]. By analysing the channel conductance of a series of substitution mutants of the phenylalanine residue, the authors confirmed that the phenyl group blocks the pore in both bacterial and plant proteins. They also obtained crystal structures of the TehA mutants, and found that the structures were consistent with the measured variations in channel conductance. Structural and functional studies thus point to a crucial role for the pore-occluding phenyl group in a previously undiscovered gating mechanism common to plant and bacterial channels.

Chen and colleagues' report opens up many fields of investigation, such as the mechanism of SLAC1 activation. The authors propose that SLAC1 phosphorylation by OST1 might induce shifts in the orientations of the channel's pore helices, resulting in the 'unlatching' of the phenyl group that blocks the pore. On the basis of previous studies of ammonium transporters[8], it is also tempting to speculate that the trimeric structure of SLAC1 could allow cooperative gating between subunits

upon phosphorylation of residues located at the interface between the subunits. Resolving the structure of a phosphorylated SLAC channel in its activated configuration would obviously be a major step in understanding the activation mechanism.

Finally, the identification of yeast Mae1 protein as an anion transporter[9] (a protein that actively moves substrates through the membrane), rather than a channel (which allows passive diffusion of the substrates), indicates that the design of the SLAC proteins, just like that of the CLC proteins[10], should allow both active and passive modes of transporting ions. Understanding the structural features that dictate whether a SLAC protein functions as a channel or as a transporter is another major goal for future studies. ∎

Sébastien Thomine and **Hélène Barbier-Brygoo** are at the Institut des Sciences du Végétal, CNRS, Gif sur Yvette 91198, France.
e-mail: Helene.Barbier-Brygoo@isv.cnrs-gif.fr

1. Chen, Y. *et al. Nature* **467,** 1074–1080 (2010).
2. Kim, T. H., Böhmer, M., Hu, H., Nishimura, N. & Schroeder, J. I. *Annu. Rev. Plant Biol.* **61,** 561–591 (2010).
3. Vahisalu, T. *et al. Nature* **452,** 487–491 (2008).
4. Geiger, D. *et al. Proc. Natl Acad. Sci. USA* **106,** 21425–21430 (2009).
5. Hille, B. *Ion Channels of Excitable Membranes* 3rd edn (Sinauer, 2001).
6. Dutzler, R., Campbell, E. B., Cadene, M., Chait, B. T. & MacKinnon, R. *Nature* **415,** 287–294 (2002).
7. De Angeli, A. *et al. Nature* **442,** 939–942 (2006).
8. Loqué, D. *et al. Nature* **446,** 195–198 (2007).
9. Camarasa, C. *et al. Appl. Environ. Microbiol.* **61,** 4144–4151 (2001).
10. Accardi, A. & Miller, C. *Nature* **427,** 803–807 (2004).

ARTICLE

A map of human genome variation from population-scale sequencing

The 1000 Genomes Project Consortium*

The 1000 Genomes Project aims to provide a deep characterization of human genome sequence variation as a foundation for investigating the relationship between genotype and phenotype. Here we present results of the pilot phase of the project, designed to develop and compare different strategies for genome-wide sequencing with high-throughput platforms. We undertook three projects: low-coverage whole-genome sequencing of 179 individuals from four populations; high-coverage sequencing of two mother–father–child trios; and exon-targeted sequencing of 697 individuals from seven populations. We describe the location, allele frequency and local haplotype structure of approximately 15 million single nucleotide polymorphisms, 1 million short insertions and deletions, and 20,000 structural variants, most of which were previously undescribed. We show that, because we have catalogued the vast majority of common variation, over 95% of the currently accessible variants found in any individual are present in this data set. On average, each person is found to carry approximately 250 to 300 loss-of-function variants in annotated genes and 50 to 100 variants previously implicated in inherited disorders. We demonstrate how these results can be used to inform association and functional studies. From the two trios, we directly estimate the rate of *de novo* germline base substitution mutations to be approximately 10^{-8} per base pair per generation. We explore the data with regard to signatures of natural selection, and identify a marked reduction of genetic variation in the neighbourhood of genes, due to selection at linked sites. These methods and public data will support the next phase of human genetic research.

Understanding the relationship between genotype and phenotype is one of the central goals in biology and medicine. The reference human genome sequence[1] provides a foundation for the study of human genetics, but systematic investigation of human variation requires full knowledge of DNA sequence variation across the entire spectrum of allele frequencies and types of DNA differences. Substantial progress has already been made. By 2008 the public catalogue of variant sites (dbSNP 129) contained approximately 11 million single nucleotide polymorphisms (SNPs) and 3 million short insertions and deletions (indels)[2–4]. Databases of structural variants (for example, dbVAR) indexed the locations of large genomic variants. The International HapMap Project catalogued both allele frequencies and the correlation patterns between nearby variants, a phenomenon known as linkage disequilibrium (LD), across several populations for 3.5 million SNPs[3,4].

These resources have driven disease gene discovery in the first generation of genome-wide association studies (GWAS), wherein genotypes at several hundred thousand variant sites, combined with the knowledge of LD structure, allow the vast majority of common variants (here, those with >5% minor allele frequency (MAF)) to be tested for association[4] with disease. Over the past 5 years association studies have identified more than a thousand genomic regions associated with disease susceptibility and other common traits[5]. Genome-wide collections of both common and rare structural variants have similarly been tested for association with disease[6].

Despite these successes, much work is still needed to achieve a deep understanding of the genetic contribution to human phenotypes[7]. Once a region has been identified as harbouring a risk locus, detailed study of all genetic variants in the locus is required to discover the causal variant(s), to quantify their contribution to disease susceptibility, and to elucidate their roles in functional pathways. Low-frequency and rare variants (here defined as 0.5% to 5% MAF, and below 0.5% MAF, respectively) vastly outnumber common variants and also contribute significantly to the genetic architecture of disease, but it has not yet been possible to study them systematically[7–9]. Meanwhile, advances in DNA sequencing technology have enabled the sequencing of individual genomes[10–13], illuminating the gaps in the first generation of databases that contain mostly common variant sites. A much more complete catalogue of human DNA variation is a prerequisite to understand fully the role of common and low-frequency variants in human phenotypic variation.

The aim of the 1000 Genomes Project is to discover, genotype and provide accurate haplotype information on all forms of human DNA polymorphism in multiple human populations. Specifically, the goal is to characterize over 95% of variants that are in genomic regions accessible to current high-throughput sequencing technologies and that have allele frequency of 1% or higher (the classical definition of polymorphism) in each of five major population groups (populations in or with ancestry from Europe, East Asia, South Asia, West Africa and the Americas). Because functional alleles are often found in coding regions and have reduced allele frequencies, lower frequency alleles (down towards 0.1%) will also be catalogued in such regions.

Here we report the results of the pilot phase of the project, the aim of which was to develop and compare different strategies for genome-wide sequencing with high-throughput platforms. To this end we undertook three projects: low-coverage sequencing of 179 individuals; deep sequencing of six individuals in two trios; and exon sequencing of 8,140 exons in 697 individuals (Box 1). The results give us a much deeper, more uniform picture of human genetic variation than was previously available, providing new insights into the landscapes of functional variation, genetic association and natural selection in humans.

Data generation, alignment and variant discovery

A total of 4.9 terabases of DNA sequence was generated in nine sequencing centres using three sequencing technologies, from DNA

obtained from immortalized lymphoblastoid cell lines (Table 1 and Supplementary Table 1). All sequenced individuals provided informed consent and explicitly agreed to public dissemination of their variation

The 1000 Genomes pilot projects

To develop and assess multiple strategies to detect and genotype variants of various types and frequencies using high-throughput sequencing, we carried out three projects, using samples from the extended HapMap collection[17].

Trio project: whole-genome shotgun sequencing at high coverage (average 42×) of two families (one Yoruba from Ibadan, Nigeria (YRI); one of European ancestry in Utah (CEU)), each including two parents and one daughter. Each of the offspring was sequenced using three platforms and by multiple centres.

Low-coverage project: whole-genome shotgun sequencing at low coverage (2–6×) of 59 unrelated individuals from YRI, 60 unrelated individuals from CEU, 30 unrelated Han Chinese individuals in Beijing (CHB) and 30 unrelated Japanese individuals in Tokyo (JPT).

Exon project: targeted capture of 8,140 exons from 906 randomly selected genes (total of 1.4 Mb) followed by sequencing at high coverage (average >50×) in 697 individuals from 7 populations of African (YRI, Luhya in Webuye, Kenya (LWK)), European (CEU, Toscani in Italia (TSI)) and East Asian (CHB, JPT, Chinese in Denver, Colorado (CHD)) ancestry.
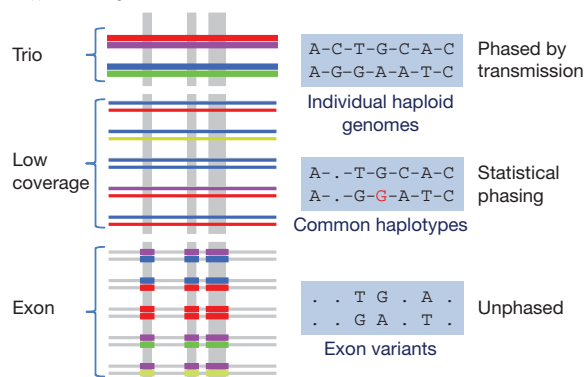
The three experimental designs differ substantially both in their ability to obtain data for variants of different types and frequencies and in the analytical methods we used to infer individual genotypes. Box 1 Figure shows a schematic representation of the projects and the type of information obtained from each. Colours in the left region indicate different haplotypes in individual genomes, and line width indicates depth of coverage (not to scale). The shaded region to the right gives an example of genotype data that could be generated for the same sample under the three strategies (dots indicate missing data; dashes indicate phase information, that is, whether heterozygous variants can be assigned to the correct haplotype). Within a short region of the genome, each individual carries two haplotypes, typically shared by others in the population. In the trio design, high-sequence coverage and the use of multiple platforms enable accurate discovery of multiple variant types across most of the genome, with Mendelian transmission aiding genotype estimation, inference of haplotypes and quality control. The low-coverage project, in contrast, efficiently identifies shared variants on common haplotypes[49,50] (red or blue), but has lower power to detect rare haplotypes (light green) and associated variants (indicated by the missing alleles), and will give some inaccurate genotypes (indicated by the red allele incorrectly assigned G). The exon design enables accurate discovery of common, rare and low-frequency variation in the targeted portion of the genome, but lacks the ability to observe variants outside the targeted regions or assign haplotype phase.

data, as part of the HapMap Project (see Supplementary Information for details of informed consent and data release). The heterogeneity of the sequence data (read lengths from 25 to several hundred base pairs (bp); single and paired end) reflects the diversity and rapid evolution of the underlying technologies during the project. All primary sequence data were confirmed to have come from the correct individual by comparison to HapMap SNP genotype data.

Analysis to detect and genotype sequence variants differed among variant types and the three projects, but all workflows shared the following four features. (1) Discovery: alignment of sequence reads to the reference genome and identification of candidate sites or regions at which one or more samples differ from the reference sequence; (2) filtering: use of quality control measures to remove candidate sites that were probably false positives; (3) genotyping: estimation of the alleles present in each individual at variant sites or regions; (4) validation: assaying a subset of newly discovered variants using an independent technology, enabling the estimation of the false discovery rate (FDR). Independent data sources were used to estimate the accuracy of inferred genotypes.

All primary sequence reads, mapped reads, variant calls, inferred genotypes, estimated haplotypes and new independent validation data are publicly available through the project website (http://www.1000genomes.org); filtered sets of variants, allele frequencies and genotypes were also deposited in dbSNP (http://www.ncbi.nlm.nih.gov/snp).

Alignment and the 'accessible genome'

Sequencing reads were aligned to the NCBI36 reference genome (details in Supplementary Information) and made available in the BAM file format[14], an early innovation of the project for storing and sharing high-throughput sequencing data. Accurate identification of genetic variation depends on alignment of the sequence data to the correct genomic location. We restricted most variant calling to the 'accessible genome', defined as that portion of the reference sequence that remains after excluding regions with many ambiguously placed reads or unexpectedly high or low numbers of aligned reads (Supplementary Information). This approach balances the need to reduce incorrect alignments and false-positive detection of variants against maximizing the proportion of the genome that can be interrogated.

For the low-coverage analysis, the accessible genome contains approximately 85% of the reference sequence and 93% of the coding sequences. Over 99% of sites genotyped in the second generation haplotype map (HapMap II)[4] are included. Of inaccessible sites, over 97% are annotated as high-copy repeats or segmental duplications. However, only one-quarter of previously discovered repeats and segmental duplications were inaccessible (Supplementary Table 2). Much of the data for the trio project were collected before technical improvements in our ability to map sequence reads robustly to some of the repeated regions of the genome (primarily longer, paired reads). For these reasons, stringent alignment was more difficult and a smaller portion of the genome was accessible in the trio project: 80% of the reference, 85% of coding sequence and 97% of HapMap II sites (Table 1).

Calibration, local realignment and assembly

The quality of variant calls is influenced by many factors including the quantification of base-calling error rates in sequence reads, the accuracy of local read alignment and the method by which individual genotypes are defined. The project introduced key innovations in each of these areas (see Supplementary Information). First, base quality scores reported by the image processing software were empirically recalibrated by tallying the proportion that mismatched the reference sequence (at non-dbSNP sites) as a function of the reported quality score, position in read and other characteristics. Second, at potential variant sites, local realignment of all reads was performed jointly across all samples, allowing for alternative alleles that contained indels. This realignment step substantially reduced errors, because local misalignment, particularly around indels, can be a major source of error in

Table 1 | Variants discovered by project, type, population and novelty

a Summary of project data including combined exon populations

Statistic	Low coverage				Trios			Exon (total)	Union across projects
	CEU	YRI	CHB+JPT	Total	CEU	YRI	Total		
Samples	60	59	60	179	3	3	6	697	742
Total raw bases (Gb)	1,402	874	596	2,872	560	615	1,175	845	4,892
Total mapped bases (Gb)	817	596	468	1,881	369	342	711	56	2,648
Mean mapped depth ($\times$)	4.62	3.42	2.65	3.56	43.14	40.05	41.60	55.92	NA
Bases accessed (% of genome)	2.43 Gb (86%)	2.39 Gb (85%)	2.41 Gb (85%)	2.42 Gb (86.0%)	2.26 Gb (79%)	2.21 Gb (78%)	2.24 Gb (79%)	1.4 Mb	NA
No. of SNPs (% novel)	7,943,827 (33%)	10,938,130 (47%)	6,273,441 (28%)	14,894,361 (54%)	3,646,764 (11%)	4,502,439 (23%)	5,907,699 (24%)	12,758 (70%)	15,275,256 (55%)
Mean variant SNP sites per individual	2,918,623	3,335,795	2,810,573	3,019,909	2,741,276	3,261,036	3,001,156	763	NA
No. of indels (% novel)	728,075 (39%)	941,567 (52%)	666,639 (39%)	1,330,158 (57%)	411,611 (25%)	502,462 (37%)	682,148 (38%)	96 (74%)	1,480,877 (57%)
Mean variant indel sites per individual	354,767	383,200	347,400	361,669	322,078	382,869	352,474	3	NA
No. of deletions (% novel)	ND	ND	ND	15,893 (60%)	6,593 (41%)	8,129 (50%)	11,248 (51%)	ND	22,025 (61%)
No. of genotyped deletions (% novel)	ND	ND	ND	10,742 (57%)	ND	ND	6,317 (48%)	ND	13,826 (58%)
No. of duplications (% novel)	259 (90%)	320 (90%)	280 (91%)	407 (89%)	187 (93%)	192 (91%)	256 (92%)	ND	501 (89%)
No. of mobile element insertions (% novel)	3,202 (79%)	3,105 (84%)	1,952 (76%)	4,775 (86%)	1,397 (68%)	1,846 (78%)	2,531 (78%)	ND	5,370 (87%)
No. of novel sequence insertions (% novel)	ND	ND	ND	ND	111 (96%)	66 (86%)	174 (93%)	ND	174 (93%)

b Exon populations separately

Statistic	CEU	TSI	LWK	YRI	CHB	CHD	JPT
Samples	90	66	108	112	109	107	105
Total collected bases (Gb)	151	64	53	147	93	127	211
Mean mapped depth on target ($\times$)	73	71	32	62	47	62	53
No. of SNPs (% novel)	3,489 (34%)	3,281 (34%)	5,459 (50%)	5,175 (46%)	3,415 (47%)	3,431 (50%)	2,900 (42%)
Variant SNP sites per individual	715	727	902	794	713	770	694
No. of indels (no. novel)	23 (10)	22 (11)	24 (16)	38 (21)	30 (16)	26 (13)	25 (11)
Variant indel sites per individual	3	3	3	3	3	2	3

NA, not applicable; ND, not determined.

variant calling. Finally, by initially analysing the data with multiple genotype and variant calling algorithms and then generating a consensus of these results, the project reduced genotyping error rates by 30–50% compared to those currently achievable using any one of the methods alone (Supplementary Fig. 1 and Supplementary Table 12).

We also used local realignment to generate candidate alternative haplotypes in the process of calling short (1–50-bp) indels[15], as well as local *de novo* assembly to resolve breakpoints for deletions greater than 50 bp. The latter resulted in a doubling of the number of large (>1 kb) structural variants delineated with base-pair resolution[16]. Full genome *de novo* assembly was also performed (Supplementary Information), resulting in the identification of 3.7 megabases (Mb) of novel sequence not matching the reference at a high threshold for assembly quality and novelty. All novel sequence matched other human and great ape sequences in the public databases.

Rates of variant discovery
In the trio project, with an average mapped sequence coverage of 42$\times$ per individual across six individuals and 2.3 gigabases (Gb) of accessible genome, we identified 5.9 million SNPs, 650,000 short indels (of 1–50 bp in length), and over 14,000 larger structural variants. In the low-coverage project, with average mapped coverage of 3.6$\times$ per individual across 179 individuals (Supplementary Fig. 2) and 2.4 Gb of accessible genome, we identified 14.4 million SNPs, 1.3 million short indels and over 20,000 larger structural variants. In the exon project, with an average mapped sequence coverage of 56$\times$ per individual across 697 individuals and a target of 1.4 Mb, we identified 12,758 SNPs and 96 indels.

Experimental validation was used to estimate and control the FDR for novel variants (Supplementary Table 3). The FDR for each complete call set was controlled to be less than 5% for SNPs and short indels, and less than 10% for structural variants. Because in an initial test

almost all of the sites that we called that were already in dbSNP were validated (285 out of 286), in most subsequent validation experiments we tested only novel variants and extrapolated to obtain the overall FDR. This process will underestimate the true FDR if more SNPs listed in dbSNP are false positives for some call sets. The FDR for novel variants was 2.6% for trio SNPs, 10.9% for low-coverage SNPs, and 1.7% for low-coverage indels (Supplementary Information and Supplementary Tables 3 and 4a, b).

Variation detected by the project is not evenly distributed across the genome: certain regions, such as the human leukocyte antigen (HLA) and subtelomeric regions, show high rates of variation, whereas others, for example a 5-Mb gene-dense and highly conserved region around 3p21, show very low levels of variation (Supplementary Fig. 3a). At the chromosomal scale we see strong correlation between different forms of variation, particularly between SNPs and indels (Supplementary Fig. 3b). However, we also find heterogeneity particular to types of structural variant, for example structural variants resulting from non-allelic homologous recombination are apparently enriched in the HLA and subtelomeric regions (Supplementary Fig. 3b, top).

Variant novelty
As expected, the vast majority of sites variant in any given individual were already present in dbSNP; the proportion newly discovered differed substantially among populations, variant types and allele frequencies (Fig. 1). Novel SNPs had a strong tendency to be found only in one analysis panel (set of related populations; Fig. 1a). For SNPs also present in dbSNP version 129 (the last release before 1000 Genomes Project data), only 25% were specific to a single low-coverage analysis panel and 56% were found in all panels. On the other hand, 84% of newly discovered SNPs were specific to a single analysis panel whereas only 4% were found in all analysis panels. In the exon project,
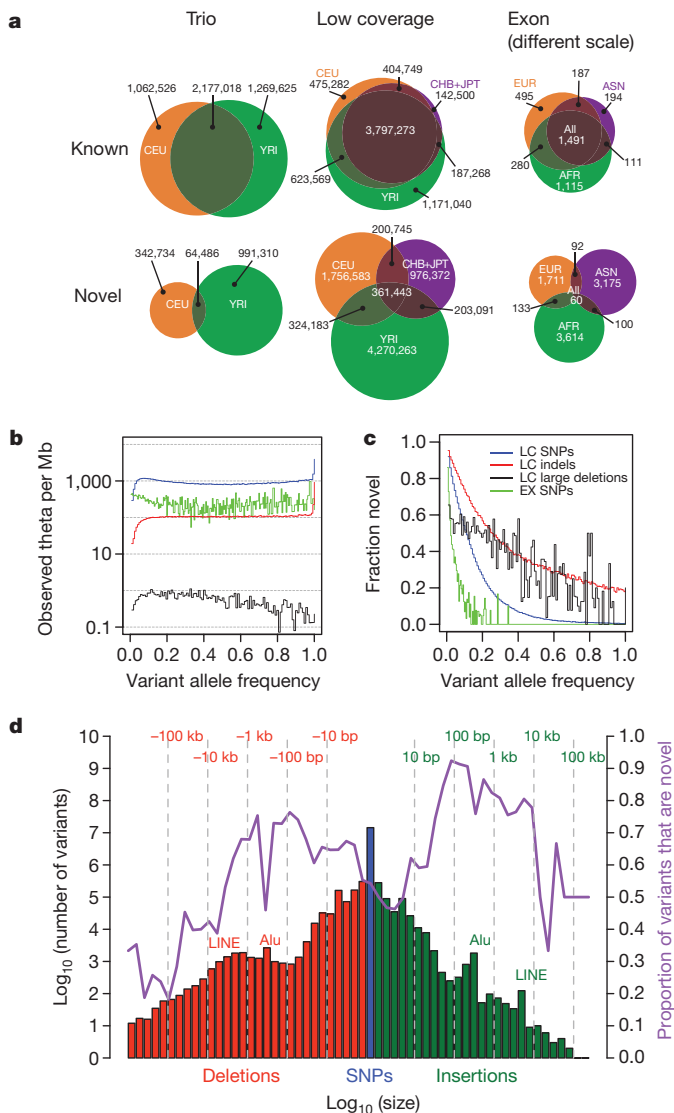
Figure 1 | Properties of the variants found. a, Venn diagrams showing the numbers of SNPs identified in each pilot project in each population or analysis panel, subdivided according to whether the SNP was present in dbSNP release 129 (Known) or not (Novel). Exon analysis panel AFR is YRI+LWK, ASN is CHB+CHD+JPT, and EUR is CEU+TSI. Note that the scale for the exon project column is much larger than for the other pilots. **b**, The number of variants per megabase (Mb) at different allele frequencies divided by the expectation under the neutral coalescent ($1/i$, where i is the variant allele count), thus giving an estimate of theta per megabase. Blue, low-coverage SNPs; red, low-coverage indels; black, low-coverage genotyped large deletions; green, exon SNPs. The spikes at the right ends of the lines correspond to excess variants for which all samples differed from the reference (approximately 1 per 30 kb), consistent with errors in the reference sequence. **c**, Fraction of variants in each allele frequency class that were novel. Novelty was determined by comparison to dbSNP release 129 for SNPs and small indels, dbVar (June 2010) for deletions, and two published genomes[10,11] for larger indels. LC, low-coverage; EX, exon. **d**, Size distribution and novelty of variants discovered in the low-coverage project. SNPs are shown in blue, deletions with respect to the reference sequence in red, and insertions or duplications with respect to the reference in green. The fraction of variants in each size bin that were novel is shown by the purple line, and is defined relative to dbSNP (SNPs and indels), dbVar (deletions, duplications, mobile element insertions), dbRIP and other studies[47] (mobile element insertions), J. C. Venter and J. Watson genomes[10,11] (short indels and large deletions), and short indels from split capillary reads[48]. To account for ambiguous placement of many indels, discovered indels were deemed to match known indels if they were within 25 bp of a known indel of the same size. To account for imprecise knowledge of the location of most deletions and duplications, discovered variants were deemed to match known variants if they had >50% reciprocal overlap.

where increased depth of coverage and sample size resulted in a higher fraction of low-frequency variants among discovered sites, 96% of novel variants were restricted to samples from a single analysis panel. In contrast, many novel structural variants were identified in all analysis panels, reflecting the lower degree of previous characterization (Supplementary Fig. 4).

Populations with African ancestry contributed the largest number of variants and contained the highest fraction of novel variants, reflecting the greater diversity in African populations. For example, 63% of novel SNPs in the low-coverage project and 44% in the exon project were discovered in the African populations, compared to 33% and 22% in the European ancestry populations.

The larger sample sizes in the exon and low-coverage projects allowed us to detect a large number of low-frequency variants (MAF <5%, Fig. 1b). Compared to the distribution expected from population genetic theory (the neutral coalescent with constant population size), we saw an excess of lower frequency variants in the exon project, reflecting purifying selection against weakly deleterious mutations and recent population growth. There are signs of a similar excess in the low-coverage project SNPs, truncated below 5% variant allele frequency by reduction in power of our call set to discover variants in this range, as discussed below.

As expected, nearly all of the high-frequency SNPs discovered here were already present in dbSNP; this was particularly true in coding regions (Fig. 1c). The public databases were much less complete for SNPs at low frequencies, for short indels and for structural variants (Fig. 1d). For example, in contrast to coding SNPs (91% of common coding SNPs described here were already present in dbSNP), approximately 50% of common short indels observed in this project were novel. These results are expected given the sample sizes used in the sequencing efforts that discovered most of the SNPs previously in dbSNP, and the more limited, and lower resolution, efforts to characterize indels and larger structural variation across the genome.

The number of structural variants that we observed declined rapidly with increasing variant length (Fig. 1d), with notable peaks corresponding to Alus and long interspersed nuclear elements (LINEs). The proportion of larger structural variants that was novel depended markedly on allele size, with variants 10 bp to 5 kb in size most likely to be novel (Fig. 1d). This is expected, as large (>5 kb) deletions and duplications were previously discovered using array-based approaches[17,18], whereas smaller structural variants (apart from polymorphic Alu insertions) had been less well ascertained before this study.

Mitochondrial and Y chromosome sequences

Deep coverage of the mitochondrial genome allowed us to manually curate sequences for 163 samples (Supplementary Information). Although variants that were fixed within an individual were consistent with the known phylogeny of the mitochondrial genome (Supplementary Fig. 5), we found a considerable amount of variation within individuals (heteroplasmy). For example, length heteroplasmy was detected in 79% of individuals compared with 52% using capillary sequencing[19], largely in the control region (Supplementary Fig. 6a). Base-substitution heteroplasmy was observed in 45% of samples, seven times higher than reported in the control region alone[19], and was spread throughout the molecule (Supplementary Fig. 6b). The extent to which this heteroplasmy arose in cell culture remains unknown, but appears low (Supplementary Information).

The Y chromosome was sequenced at an average depth of 1.8× in the 77 males in the low-coverage project, and 15.2× depth in the two trio fathers. Using customized analysis methods (Supplementary Information), we identified 2,870 variable sites, 74% novel, with 55 out of 56 passing independent validation. The Y chromosome phylogeny derived from the new variants identified novel, well supported clades within some of the 12 major haplogroups represented among the samples (for example, O2b in China and Japan; Supplementary Fig. 7). A striking pattern indicative of a recent rapid expansion

specific to haplogroup R1b was observed, consistent with the postulated Neolithic origin of this haplogroup in Europe[20].

Power to detect variants

The ability of sequencing to detect a site that is segregating in the population is dominated by two factors: whether the non-reference allele is present among the individuals chosen for sequencing, and the number of high-quality and well-mapped reads that overlap the variant site in individuals who carry it. Simple models show that for a given total amount of sequencing, the number of variants discovered is maximized by sequencing many samples at low coverage[21,22]. This is because high coverage of a few genomes, although providing the highest sensitivity and accuracy in genotyping a single individual, involves considerable redundancy and misses variation not represented by those samples. The low-coverage project provides us with an empirical view of the power of low-coverage sequencing to detect variants of different types and frequencies.

Figure 2a shows the rate of discovery of variants in the CEU (see Box 1 for definitions of this and other populations) samples of the low-coverage project as assessed by comparison to external data sources: HapMap and the exon project for SNPs and array CGH data[18] for large deletions. We estimate that although the low-coverage project had only ~25% power to detect singleton SNPs, power to detect SNPs present five times in the 120 sampled chromosomes was ~90% (depending on the comparator), and power was essentially complete for those present ten or more times. Similar results were seen in the YRI and CHB+JPT analysis panels at high allele counts, but slightly worse performance for variants present five times (~85% and 75%, respectively, at HapMap II sites; Supplementary Fig. 8). These results indicate that SNP discovery is less affected by the extent of LD (which is lowest in the YRI) than by sequencing coverage (which was lowest in the CHB and JPT panels).

For deletions larger than 500 bp, power was approximately 40% for singletons and reached 90% for variants present ten times or more in the sample set. Our use of several algorithms for structural variant discovery ensured that all major mechanistic subclasses of deletions were found in our analyses (Supplementary Fig. 9). The lack of appropriate comparator data sets for short indels and larger structural variants other than deletions prevented a detailed assessment of the power to detect these types of variants. However, power to detect short indels was approximately 70% for variants present at least five times in the sample, based on the rediscovery of indels in samples overlapping with the SeattleSNPs project[23]. Extrapolating from comparisons to Alu insertions discovered in the J. C. Venter genome[24] indicated an average sensitivity for common mobile element insertions of about 75%. Analysis of a set of duplications[18] indicated that only 30–40% of common duplications were discovered here, mostly as deletions with respect to the reference. Methods capable of discovering inversions and novel sequence insertions in low-coverage data with comparable specificity remain to be developed.

In summary, low-coverage shotgun sequencing provided modest power for singletons in each sample (~25–40%), and very good power for variants seen five or more times in the samples sequenced. We estimate that there was approximately 95% power to find SNPs with 5% allele frequency in the sequenced samples, and nearly 90% power to find SNPs with 5% allele frequency in populations related by 1% divergence (Fig. 2b). Thus, we believe that the projects found almost all accessible common variation in the sequenced populations and the vast majority of common variants in closely related populations.

Genotype accuracy

Genotypes, and, where possible, haplotypes, were inferred for most variants in each project (see Supplementary Information and Table 1). For the low-coverage data, statistically phased SNP genotypes were derived by using LD structure in addition to sequence information at each site, in part guided by the HapMap 3 phased haplotypes. SNP
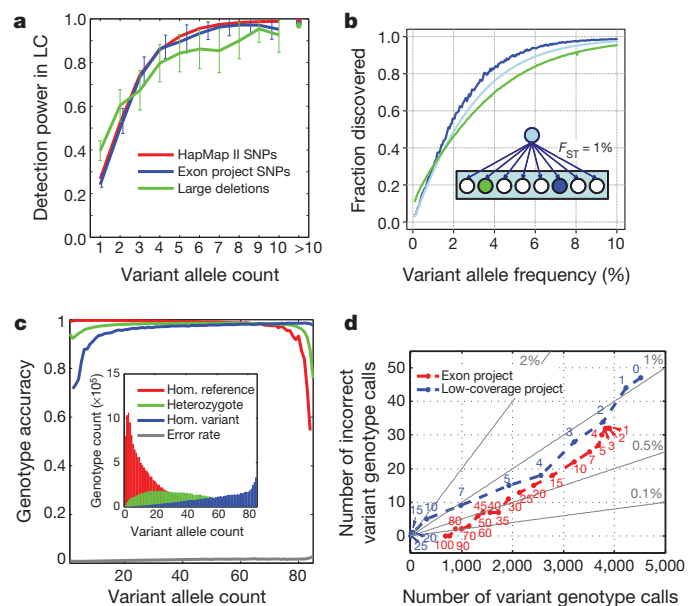
Figure 2 | Variant discovery rates and genotype accuracy in the low-coverage project. a, Rates of low-coverage variant detection by allele frequency in CEU. Lines show the fraction of variants seen in overlapping samples in independent studies that were also found to be polymorphic in the low-coverage project (in the same overlapping samples), as a function of allele count in the 60 low-coverage samples. Note that we plot power against expected allele count in 60 samples; for example, a variant present in, say, 2 copies in an overlap of 30 samples is expected to be present 4 times in 60 samples. The crosses on the right represent the average discovery fraction for all variants having more than 10 copies in the sample. Red, HapMap II sites, excluding sites also in HapMap 3 (43 overlapping samples); blue, exon project sites (57 overlapping samples); green, deletions from ref. 18 (60 overlapping samples; deletions were classified as 'found' if there was any overlap). Error bars show 95% confidence interval. **b,** Estimated rates of discovery of variants at different frequencies in the CEU (blue), a population related to the CEU with $F_{st} = 1\%$ (green), and across Europe as a whole (light blue). Inset: cartoon of the statistical model for population history and thus allele frequencies in related populations where an ancestral population gave rise to many equally related populations, one of which (blue circle) has samples sequenced. **c,** SNP genotype accuracy by allele frequency in the CEU low-coverage project, measured by comparison to HapMap II genotypes at sites present in both call sets, excluding sites that were also in HapMap 3. Lines represent the average accuracy of homozygote reference (red), heterozygote (green) and homozygote alternative calls (blue) as a function of the alternative allele count in the overlapping set of 43 samples, and the overall genotype error rate (grey, at bottom of plot). Inset: number of each genotype class as a function of alternative allele count. **d,** Coverage and accuracy for the low-coverage and exon projects as a function of depth threshold. For 41 CEU samples sequenced in both the exon and low-coverage projects, on the x axis is shown the number of non-reference SNP genotype calls at HapMap II sites not in HapMap 3 that were called in the exon project target region, and on the y axis is shown the number of these calls that were not variant (that is, are reference homozygote and thus incorrectly were called as variant) according to HapMap II. Each point plotted corresponds to a minimum depth threshold for called sites. Grey lines show constant error rates. The exon project calls (red) were made independently per sample, whereas the low-coverage calls (blue), which were only slightly less accurate, were made using LD information that combined partial information across samples and sites in an imputation-based algorithm. The additional data added from point '1' to point '0' (upper right in the figure) for the low-coverage project were completely imputed.

genotype accuracy varied considerably between projects (trio, low coverage and exon), and as a function of coverage and allele frequency. In the low-coverage project, the overall genotype error rate (based on a consensus of multiple methods) was 1–3% (Fig. 2c and Supplementary Fig. 10). The use of HapMap 3 data greatly assisted phasing of the CEU and YRI samples, for which the HapMap 3 genotypes were phased by transmission, but had a more modest effect on

genotype accuracy away from HapMap 3 sites (for further details see Supplementary Information).

The accuracy at heterozygous sites, a more sensitive measure than overall accuracy, was approximately 90% for the lowest frequency variants, increased to over 95% for intermediate frequencies, and dropped to 70–80% for the highest frequency variants (that is, those where the reference allele is the rare allele). We note that these numbers are derived from sites that can be genotyped using array technology, and performance may be lower in harder to access regions of the genome. We find only minor differences in genotype accuracy between populations, reflecting differences in coverage as well as haplotype diversity and extent of LD.

The accuracy of genotypes for large deletions was assessed against previous array-based analyses[18] (Supplementary Fig. 11). The genotype error rate across all allele frequencies and genotypes was <1%, with the accuracy of heterozygous genotypes at low (MAF <3%), intermediate (MAF ~50%) and high-frequency (MAF >97%) variants estimated at 86%, 97% and 83%, respectively. The greater apparent genotype accuracy of structural variants compared to SNPs in the low-coverage project reflects the increased number of informative reads per individual for variants of large size and a bias in the known large deletion genotype set for larger, easier to genotype variants.

For calling genotypes in the low-coverage samples, the utility of using LD information in addition to sequence data at each site was demonstrated by comparison to genotypes of the exon project, which were derived independently for each site using high-coverage data. Figure 2d shows the SNP genotype error rate as a function of depth at the genotyped sites in CEU. A similar number of variants was called, and at comparable accuracy, using minimum 4× depth in the low-coverage project as was obtained with minimum 15× depth in the exon project. To genotype a high fraction of sites both projects needed to make calls at sites with low coverage, and the LD-based calling strategy for the low-coverage project used imputation to make calls at nearly 15% more sites with only a modest increase in error rate.

The accuracy and completeness of the individual genome sequences in the low-coverage project could be estimated from the trio mothers, each of whom was sequenced to high coverage, and for whom data subsampled to 4× were included in the low-coverage analysis. Comparison of the SNP genotypes in the two projects showed that where the CEU mother had at least one variant allele according to the trio analysis, in 96.9% of cases the variant was also identified in the low-coverage project and in 93.8% of cases the genotype was accurately inferred. For the YRI trio mother the equivalent figures are 95.0% and 88.4%, respectively (note that false positives in the trio calls will lead to underestimates of the accuracy).

Putative functional variants

An individual's genome contains many variants of functional consequence, ranging from the beneficial to the highly deleterious. We estimated that an individual typically differs from the reference human genome sequence at 10,000–11,000 non-synonymous sites (sequence differences that lead to differences in the protein sequence) in addition to 10,000–12,000 synonymous sites (differences in coding exons that do not lead to differences in the protein sequence; Table 2). We found a much smaller number of variants likely to have greater functional impact: 190–210 in-frame indels, 80–100 premature stop codons, 40–50 splice-site-disrupting variants and 220–250 deletions that shift reading frame, in each individual. We estimated that each genome is heterozygous for 50–100 variants classified by the Human Gene Mutation Database (HGMD) as causing inherited disorders (HGMD-DM). Estimates from the different pilot projects were consistent with each other, taking into consideration differences in power to detect low-frequency variants, fraction of the accessible genome and population differences (Table 2), as well as with previous observations based on personal genome sequences[10,11]. Collectively, we refer to the 340–400 premature stops, splice-site disruptions and frame shifts, affecting 250–300 genes per individual, as putative loss-of-function (LOF) variants.

In total, we found 68,300 non-synonymous SNPs, 34,161 of which were novel (Table 2). In an early analysis, 21,657 non-synonymous SNPs were validated as polymorphic in 620 samples using a custom genotyping array (Supplementary Information). The mean minor allele frequency in the array data was 2.2% for 4,573 novel variants, and 26.2% for previously discovered variants.

Overall we rediscovered 671 (1.3%) of the 50,361 coding single nucleotide variants in HGMD-DM (Supplementary Table 5). The types of disease for which variants were identified were biased towards certain categories (Supplementary Fig. 12), with diseases associated with the eye and reproduction significantly over represented and diseases of the nervous system significantly under represented. These biases reflect multiple factors including differences in the fitness effects of the variants, the extent of medical genetics research and differences in the false reporting rate among 'disease causing' variants.

As expected, and consistent with purifying selection, putative functional variants had an allele frequency spectrum depleted at higher allele frequencies, with putative LOF variants showing this effect more strongly (Supplementary Fig. 13). Of the low-coverage non-synonymous, stop-introducing, splice-disrupting and HGMD-DM variants, 67.3%, 77.3%, 82.2% and 84.7% were private to single populations, compared to 61.1% for synonymous variants. Across these same functional classes, 15.8%, 25.9%, 21.6% and 19.9% of variants were found in only a single individual, compared to 11.8% of synonymous variants.

The tendency for deleterious functional variants to have lower allele frequencies has consequences for the discovery and analysis of this type of variation. In the deeply sequenced CEU trio father, who was not included in the low-coverage project, 97.8% of all single base variants had been found in the low-coverage project, but only 95% of non-synonymous, 88% of stop-inducing and 85% of HGMD-DM variants. The missed variants correspond to 389 non-synonymous, 11 stop-inducing and 13 HGMD-DM variants. As sample size increases, the number of novel variants per sequenced individual will decrease, but only slowly. Analyses based on the exon project data (Fig. 3)

Table 2 | Estimated numbers of potentially functional variants in genes

Class	Combined total	Combined novel	Low coverage		High-coverage trio		Exon capture		
			Total	Interquartile*	Total	Individual range	Total	Interquartile*	GENCODE extrapolation
Synonymous SNPs	60,157	23,498	55,217	10,572–12,126	21,410	9,193–12,500	5,708	461–532	11,553–13,333
Non-synonymous SNPs	68,300	34,161	61,284	9,966–10,819	19,824	8,299–10,866	7,063	396–441	9,924–11,052
Small in-frame indels	714	383	666	198–205	289	130–178	59	1–3	~25–75
Stop losses	77	40	71	9–11	22	4–14	6	0–0	~0–0
Stop-introducing SNPs	1,057	755	951	88–101	192	67–100	82	2–3	~50–75
Splice-site-disrupting SNPs	517	399	500	41–49	82	28–45	3	1–1	~50
Small frameshift indels	954	551	890	227–242	433	192–280	37	0–1	~0–25
Genes disrupted by large deletions	147	71	143	28–36	82	33–49	ND	ND	ND
Total genes containing LOF variants	2,304	NA	1,795	272–297	483	240–345	77	3–4	~75–100
HGMD 'damaging mutation' SNPs	671	NA	578	57–80	161	48–82	99	2–4	~50–100

NA, not applicable; ND, not determined.
* Interquartile range of the number of variants of specified type per individual.
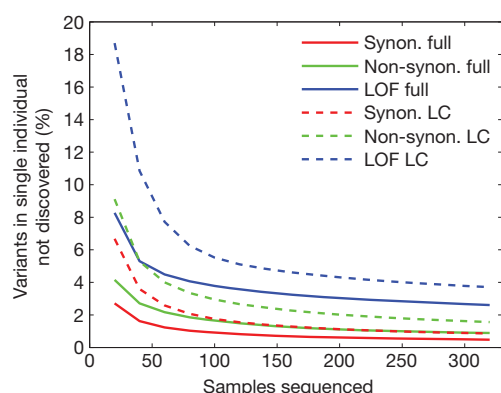
Figure 3 | The value of additional samples for variant discovery. The fraction of variants present in an individual that would not have been found in a sequenced reference panel, as a function of reference panel size and the sequencing strategy. The lines represent predictions for synonymous (Synon.), non-synonymous (Non-synon.) and loss-of-function (LOF) variant classes, broken down by sequencing category: full sequencing as for exons (Full) and low-coverage sequencing (LC). The values were calculated from observed distributions of variants of each class in 321 East Asian samples (CHB, CHD and JPT populations) in the exon data, and power to detect variants at low allele counts in the reference panel from Fig. 2a.

showed that, on average, 99% of the synonymous variants in an individual would be found in 100 deeply sequenced samples, whereas 250 samples would be required to find 99% of non-synonymous variants and 320 samples would still find only 97.4% of the LOF variants present in an individual. Using detection power data from Fig. 2a, we estimated that 250 samples sequenced at low coverage would be needed to find 99% of the synonymous variants in an individual, and with 320 sequenced samples 98.5% of non-synonymous and 96.3% of LOF variants would be found.

Application to association studies

Whole-genome sequencing enables all genetic variants present in a sample set to be tested directly for association with a given disease or trait. To quantify the benefit of having more complete ascertainment of genetic variation beyond that achievable with genotyping arrays, we carried out expression quantitative trait loci (eQTL) association tests on the 142 low-coverage samples for which expression data are available in the cell lines[25]. When association analysis (Spearman rank correlation, FDR <5%, eQTLs within 50 kb of probe) was performed using all sites discovered in the low-coverage project, a larger number of significant eQTLs (increase of ~20% to 50%) was observed as compared to association analysis restricted to sites present on the Illumina 1M chip (Supplementary Table 6). The increase was lower in the CHB+JPT and CEU samples, where greater LD exists between previously examined and newly discovered variants, and higher in the YRI samples, where there are more novel variants and less LD. These results indicate that, while modern genotyping arrays capture most of the common variation, there remain substantial additional contributions to phenotypic variation from the variants not well captured by the arrays.

Population sequencing of large phenotyped cohorts will allow direct association tests for low-frequency variants, with a resolution determined by the LD structure. An alternative that is less expensive, albeit less accurate, is to impute variants from a sequenced reference panel into previously genotyped samples[26,27]. We evaluated the accuracy of imputation that uses the current low-coverage project haplotypes as the reference panel. Specifically, we compared genotypes derived by deep sequencing of one individual in each trio (the fathers) with genotypes derived using the HapMap 3 genotype data (which combined data from the Affymetrix 6.0 and Illumina 1M arrays) in those same two individuals and imputation based on the low-coverage

project haplotypes to fill in their missing genotypes. At variant sites (that is, where the father was not homozygous for the reference sequence), imputation accuracy was highest for SNPs at which the minor allele was observed at least six times in our low-coverage samples, with an error rate of ~4% in CEU and ~10% in YRI, and became progressively worse for rarer SNPs, with error rates of 35% for sites where the minor allele was observed only twice in the low-coverage samples (Fig. 4a).

Although the ability to impute rare variants accurately from the 1000 Genomes Project resource is currently limited, the completeness of the resource nevertheless increases power to detect association signals. To demonstrate the utility of imputation in disease samples, we imputed into an eQTL study of ~400 children of European ancestry[28] using the low-coverage pilot data and HapMap II as reference panels. By comparison to directly genotyped sites we estimated that the effective sample size at variants imputed from the pilot CEU low-coverage data set is 91% of the true sample size for variants with allele frequencies above 10%, 76% in the allele frequency range 4–6%, and 54% in the range 1–2%. Imputing over 6 million variants from the low-coverage project data increased the number of detected cis-eQTLs by ~16%, compared to a 9% increase with imputing from HapMap II (FDR 5%, signal within 50 kb of transcript; for an example see Fig. 4b).

In addition to this modest increase in the number of discoveries, testing almost all common variants allows identification of many additional candidate variants that might underlie each association. For example, we find that rs11078928, a variant in a splice site for *GSDMB*, is in strong LD with SNPs near *ORMDL3*, previously associated with asthma, Crohn's disease, type 1 diabetes and rheumatoid arthritis, thus leading to the hypothesis that *GSDMB* could be the causative gene in these associations. Although rs11078928 is not newly discovered, it was not included in HapMap or on commercial SNP arrays, and thus could not have been identified as associated with
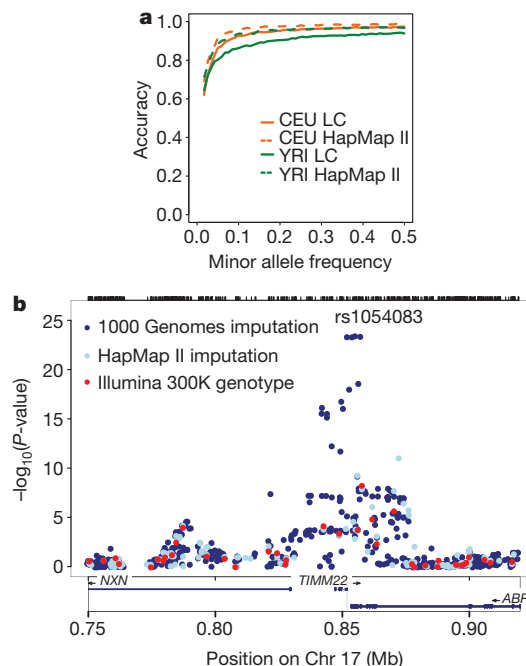
Figure 4 | Imputation from the low-coverage data. a, Accuracy of imputing variant genotypes using HapMap 3 sites to impute sites from the low-coverage (LC) project into the trio fathers as a function of allele frequency. Accuracy of imputing genotypes from the HapMap II reference panels[4] is also shown. Imputation accuracy for common variants was generally a few per cent worse from the low-coverage project than from HapMap, although error rates increase for less common variants. **b,** An example of imputation in a cis-eQTL for *TIMM22*, for which the original Ilumina 300K genotype data gave a weak signal[28]. Imputation using HapMap data made a small improvement, and imputation using low-coverage haplotypes provided a much stronger signal.

these diseases before this project. Similarly, a recent study[29] used project data to show that coding variants in *APOL1* probably underlie a major risk for kidney disease in African-Americans previously attributed (at a lower effect size) to *MYH9*. These examples demonstrate the value of having much more complete information on LD, the almost complete set of common variants, and putative functional variants in known association intervals.

Testing almost all common variants also allows us to examine general properties of genetic association signals. The NHGRI GWAS catalogue (http://www.genome.gov/gwastudies, accessed 15 July 2010) described 1,227 unique SNPs associated with one or more traits ($P < 5 \times 10^{-8}$). Of these, 1,185 (96.5%) are present in the low-coverage CEU data set. Under 30% of these are either annotated as non-synonymous variants (77, 6.5%) or in substantial LD ($r^2 > 0.5$) with a non-synonymous variant (272, 23%). In the latter group, only 93 (8.4%) are in strong LD ($r^2 > 0.9$) with a non-synonymous variant. Because we tested ~95% of common variation, these results indicate that no more than one-third of complex trait association signals are likely to be caused by common coding variation. Although it remains to be seen whether reported associations are better explained through weak LD to coding variants with strong effects, these results are consistent with the view that most contributions of common variation to complex traits are regulatory in nature.

Mutation, recombination and natural selection

Project sequence data allowed us to investigate fundamental processes that shape human genetic variation including mutation, recombination and natural selection.

Detecting *de novo* mutations in trio samples

Deep sequencing of individuals within a pedigree offers the potential to detect *de novo* germline mutation events. Our approach was to allow a relatively high FDR in an initial screen to capture a large fraction of true events and then use a second technology to rule out false-positive mutations.

In the CEU and YRI trios, respectively, 3,236 and 2,750 candidate *de novo* germline single-base mutations were selected for further study, based on their presence in the child but not the parents. Of these, 1,001 (CEU) and 669 (YRI) were validated by re-sequencing the cell line DNA. When these were tested for segregation to offspring (CEU) or in non-clonal DNA from whole blood (YRI), only 49 CEU and 35 YRI candidates were confirmed as true germline mutations. Correcting for the fraction of the genome accessible to this analysis provided an estimate of the per generation base pair mutation rate of 1.2×10^{-8} and 1.0×10^{-8} in the CEU and YRI trios, respectively. These values are similar to estimates obtained from indirect evolutionary comparisons[30], direct studies based on pathogenic mutations[31], and a recent analysis of a single family[32].

We infer that the remaining vast majority (952 CEU and 634 YRI) of the validated variants were somatic or cell line mutations. The greater number of these validated non-germline mutations in the CEU cell line perhaps reflects the greater age of the CEU cell culture. Across the two trio offspring, we observed a single, synonymous, coding germline mutation, and 17 coding non-germline mutations of which 16 were non-synonymous, perhaps indicative of selection during cell culture.

Although the number of non-germline variants found per individual is a very small fraction of the total number of variants per individual (~0.03% for the CEU child and ~0.02% for the YRI child), these variants will not be shared between samples. Assuming that the number of non-germline mutations in these two trios is representative of all cell line DNA we analysed, we estimate that non-germline mutations might constitute 0.36% and 2.4% of all variants, and 0.61% and 3.1% of functional variants, in the low-coverage and exon pilots, respectively. In larger samples, of thousands, the overall false-positive rates from cell line mutations would become significant, and confound

interpretation, indicating that large-scale studies should use DNA from primary tissue, such as blood, where possible.

The effects of selection on local variation

Natural selection can affect levels of DNA variation around genes in several ways: strongly deleterious mutations will be rapidly eliminated by natural selection, weakly deleterious mutations may segregate in populations but rarely become fixed, and selection at nearby sites (both purifying and adaptive) reduces genetic variation through background selection[33] and the hitch-hiking effect[34]. The effect of these different forces on genetic variation can be disentangled by examining patterns of diversity and divergence within and around known functional elements. The low-coverage data enables, for the first time, genome-wide analysis of such patterns in multiple populations. Figure 5a (top panel) shows the pattern of diversity relative to genic regions measured by aggregating estimates of heterozygosity around protein-coding genes. Within genes, exons harbour the least diversity (about 50% of that of introns) and 5′ and 3′ UTRs harbour slightly less diversity than immediate flanking regions and introns. However, this variation in diversity is fully explained by the level of divergence
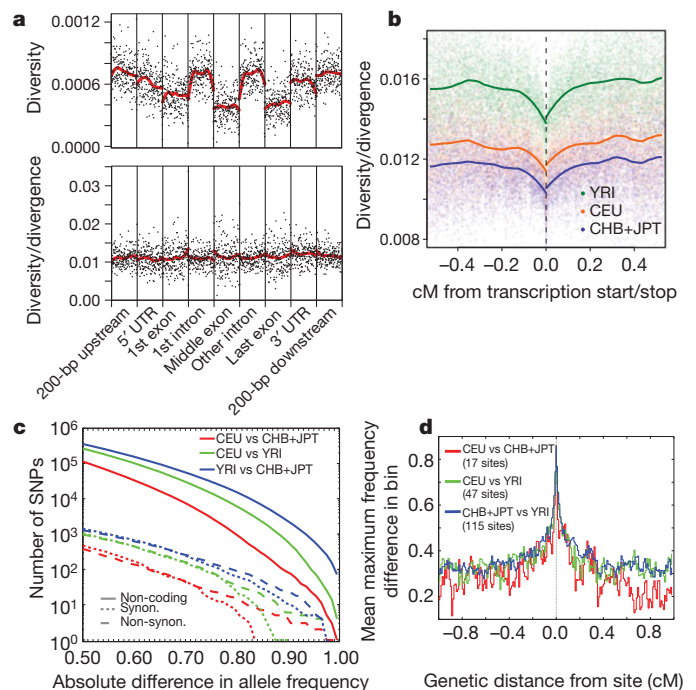
Figure 5 | Variation around genes. a, Diversity in genes calculated from the CEU low-coverage genotype calls (top) and diversity divided by divergence between humans and rhesus macaque (bottom). Within each element averaged diversity is shown for the first and last 25 bp, with the remaining 150 positions sampled at fixed distances across the element (elements shorter than 150 bp were not considered). Note that estimates of diversity will be reduced compared to the true population value due to the reduced power for rare variants, but relative values should be little affected. **b**, Average autosomal diversity divided by divergence, as a function of genetic distance from coding transcripts, calculated at putatively neutral sites, that is, excluding phastcons, conserved non-coding sequences and all sites in coding exons but fourfold degenerate sites. **c**, Numbers of SNPs showing increasingly high levels of differentiation in allele frequency between the CEU and CHB+JPT (red), CEU and YRI (green) and CHB+JPT and YRI (blue). Lines indicate synonymous variants (dashed), non-synonymous variants (dotted) and other variants (solid lines). The most highly differentiated genic SNPs were enriched for non-synonymous variants, indicating local adaptation. **d**, The decay of population differentiation around genic SNPs showing extreme allele frequency differences between populations (difference in frequency of at least 0.8 between populations, thinned so there is no more than one per gene considered; Supplementary Table 8). For all such SNPs the highest allele frequency difference in bins of 0.01 cM away from the variant was recorded and averaged.

(Fig. 5a, bottom panel), consistent with the common part of the allele frequency spectrum being dominated by effectively neutral variants, and weakly deleterious variants contributing only to the rare end of the frequency spectrum.

In contrast, diversity in the immediate vicinity of genes (scaled by divergence) is reduced by approximately 10% relative to sites distant from any gene (Fig. 5b). Although a similar reduction has been seen previously in gene-dense regions[35], project data enable the scale of the effect to be determined. We find that the reduction extends up to 0.1 cM away from genes, typically 85 kb, indicating that selection at linked sites restricts variation relative to neutral levels across the majority of the human genome.

Population differentiation and positive selection

Previous inferences about demographic history and the role of local adaptation in shaping human genetic variation made from genome-wide genotype data[4,36,37] have been limited by the partial and complex ascertainment of SNPs on genotyping arrays. Although data from the 1000 Genomes Project pilots are neither fully comprehensive nor fully free of ascertainment bias (issues include low power for rare variants, noise in allele frequency estimates, some false positives, non-random data collection across samples, platforms and populations, and the use of imputed genotypes), they can be used to address key questions about the extent of differentiation among populations, the presence of highly differentiated variants and the ability to fine-map signals of local adaptation.

Although the average level of population differentiation is low (at sites genotyped in all populations the mean value of Wright's F_{st} is 0.071 between CEU and YRI, 0.083 between YRI and CHB+JPT, and 0.052 between CHB+JPT and CEU), we find several hundred thousand SNPs with large allele frequency differences in each population comparison (Fig. 5c). As seen in previous studies[4,37], the most highly differentiated sites were enriched for non-synonymous variants, indicative of the action of local adaptation. The completeness of common variant discovery in the low-coverage resource enables new perspectives in the search for local adaptation. First, it provides a more comprehensive catalogue of fixed differences between populations, of which there are very few: two between CEU and CHB+JPT (including the A111T missense variant in *SLC24A5* (ref. 38) contributing to light skin colour), four between CEU and YRI (including the −46 GATA box null mutation upstream of *DARC*[39], the Duffy O allele leading to *Plasmodium vivax* malaria resistance) and 72 between CHB+JPT and YRI (including 24 around the exocyst complex component gene *EXOC6B*); see Supplementary Table 7 for a complete list. Second, it provides new candidates for selected variants, genes and pathways. For example, we identified 139 non-synonymous variants showing large allele frequency differences (at least 0.8) between populations (Supplementary Table 8), including at least two genes involved in meiotic recombination—*FANCA* (ninth most extreme non-synonymous SNP in CEU versus CHB+JPT) and *TEX15* (thirteenth most extreme non-synonymous SNP in CEU versus YRI, and twenty-sixth most extreme non-synonymous SNP in CHB+JPT versus YRI). Because we are finding almost all common variants in each population, these lists should contain the vast majority of the near fixed differences among these populations. Finally, it improves the fine mapping of selective sweeps (Supplementary Fig. 14) and analysis of the dynamics of location adaptation. For example, we find that the signal of population differentiation around high F_{st} genic SNPs drops by half within, on average, less than 0.05 cM (typically 30–50 kb; Fig. 5d). Furthermore, 51% of such variants are polymorphic in both populations. These observations indicate that much local adaptation has occurred by selection acting on existing variation rather than new mutation.

The effect of recombination on local sequence evolution

We estimated a fine-scale genetic map from the phased low-coverage genotypes. Recombination hotspots were narrower than previously

estimated[4] (mean hotspot width of 2.3 kb compared to 5.5 kb in HapMap II; Fig. 6a), although, unexpectedly, the estimated average peak recombination rate in hotspots is lower in YRI (13 cM Mb^{-1}) than in CEU and CHB+JPT (20 cM Mb^{-1}). In addition, crossover activity is less concentrated in the genome in YRI, with 70% of recombination occurring in 10% of the sequence rather than 80% of the recombination for CEU and CHB+JPT (Fig. 6b). A possible biological basis for these differences is that PRDM9, which binds a DNA motif strongly enriched in hotspots and influences the activity of LD-defined hotspots[40–43], shows length variation in its DNA-binding zinc fingers within populations, and substantial differentiation between African and non-African populations, with a greater allelic diversity in Africa[43]. This could mean greater diversity of hotspot locations within Africa and therefore a less concentrated picture in this data set of recombination and lower usage of LD-defined hotspots (which require evidence in at least two populations and therefore will not reflect hotspots present only in Africa).

The low-coverage data also allowed us to address a long-standing debate about whether recombination has any local mutagenic effect. Direct examination of diversity around hotspots defined from LD data are potentially biased (because the detection of hotspots requires variation to be present), but we can, without bias, examine rates of SNP variation and recombination around the PRDM9 binding motif
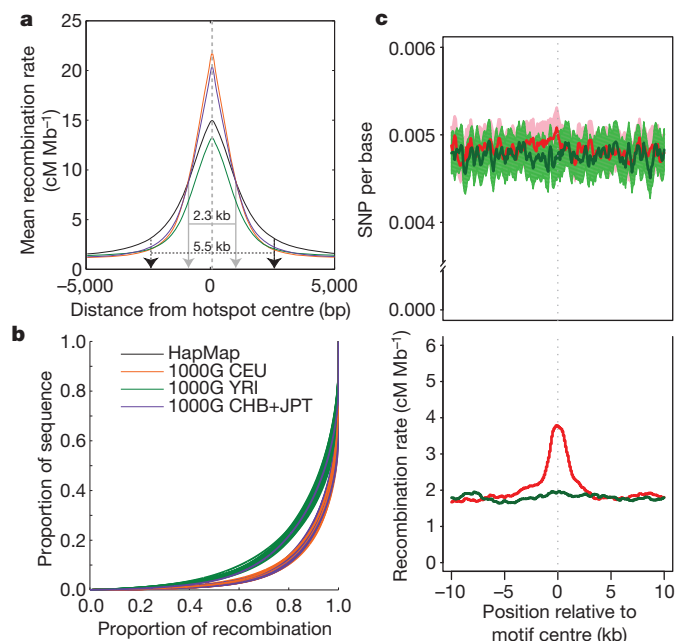
Figure 6 | Recombination. a, Improved resolution of hotspot boundaries. The average recombination rate estimated from low-coverage project data around recombination hotspots detected in HapMap II. Recombination hotspots were narrower, and in CEU (orange) and CHB+JPT (purple) more intense than previously estimated. See panel **b** for key. **b**, The concentration of recombination in a small fraction of the genome, one line per chromosome. If recombination were uniformly distributed throughout the genome, then the lines on this figure would appear along the diagonal. Instead, most recombination occurs in a small fraction of the genome. Recombination rates in YRI (green) appeared to be less concentrated in recombination hotspots than CEU (orange) or CHB+JPT (purple). HapMap II estimates are shown in black. **c**, The relationship between genetic variation and recombination rates in the YRI population. The top plot shows average levels of diversity, measured as mean number of segregating sites per base, surrounding occurrences of the previously described hotspot motif[40] (CCTCCCTNNCCAC, red line) and a closely related, but not recombinogenic, DNA sequence (CTTCCCTNNCCAC, green line). The lighter red and green shaded areas give 95% confidence intervals on diversity levels. The bottom plot shows estimated mean recombination rates surrounding motif occurrences, with colours defined as in the top plot.

BOX 2

Design of the full 1000 Genomes Project

The production phase of the full 1000 Genomes Project will combine low-coverage whole-genome sequencing, array-based genotyping, and deep targeted sequencing of all coding regions in 2,500 individuals from five large regions of the world (five population samples of 100 in or with ancestry from each of Europe, East Asia, South Asia and West Africa, and seven populations totalling 500 from the Americas; Supplementary Table 9). We will increase the low-coverage average depth to over $4\times$ per individual, and use blood-derived DNA where possible to minimize somatic and cell-line false positives.

A clustered sampling approach was chosen to improve low-frequency variant detection in comparison to a design in which a smaller number of populations was sampled to a greater depth. In a region containing a cluster of related populations, genetic drift can lead variants that are at low frequency overall to be more common (hence, easily detectable) in one population but less common (hence, likely to be undetectable) in another. We modelled this process using project data (see Supplementary Information) assuming that five sampled populations are equally closely related to each other ($F_{st} = 1\%$). We found that the low-coverage sequencing in this design would discover 95% of variants in the accessible genome at 1% frequency across each broad geographic region, between 90% and 95% of variants at 1% frequency in any one of the sampled populations, and about 85% of variants at 1% frequency in any equally related but unsampled population. Box 2 Figure shows predicted discovery curves for variants at different frequencies with details as for Fig. 2b. The model is conservative, in that it ignores migration and the contribution to discovery from more distantly related populations, each of which will increase sensitivity for variants in any given population. In exons, the full project should have 95% power to detect variants at a frequency of 0.3% and approximately 60% power for variants at a frequency of 0.1%.

In addition to improved detection power, we expect the full project to have increased genotype accuracy due to (1) advances in sequencing technology that are reducing per base error rates and alignment artefacts; (2) increased sample size, which improves imputation-based methods; (3) ongoing algorithmic improvements; and (4) the designing by the project of genotyping assays that will directly genotype up to 10 million common and low-frequency variants (SNPs, indels and structural variants) observed in the low-coverage data. In addition, we expect the fraction of the genome that is accessible to increase. Longer read lengths, improved protocols for generating paired reads, and the use of more powerful assembly and alignment methods are expected to increase accessibility from under 85% to above 90% of the reference genome (Supplementary Fig. 15).
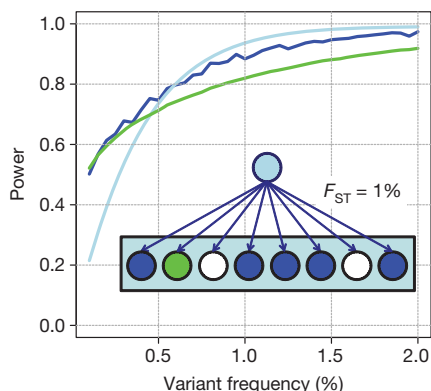
associated with hotspots. Figure 6c shows the local recombination rate and pattern of SNP variation around the motif compared to the same plots around a motif that is a single base difference away. Although the motif is associated with a sharp peak in recombination rate, there is no systematic effect on local rates of SNP variation. We infer that, although recombination may influence the fate of new mutations, for example through biased gene conversion, there is no evidence that it influences the rate at which new variants appear.

Discussion

The 1000 Genomes Project launched in 2008 with the goal of creating a public reference database for DNA polymorphism that is 95% complete at allele frequency 1%, and more complete for common variants and exonic variants, in each of multiple human population groups. The three pilot projects described here were designed to develop and evaluate methods to use high-throughput sequencing to achieve these goals. The results indicate (1) that robust protocols now exist for generating both whole-genome shotgun and targeted sequence data; (2) that algorithms to detect variants from each of these designs have been validated; and (3) that low-coverage sequencing offers an efficient approach to detect variation genome wide, whereas targeted sequencing offers an efficient approach to detect and accurately genotype rare variants in regions of functional interest (such as exons).

Data from the pilot projects are already informing medical genetic studies. As shown in our analysis of previous eQTL data sets, a more complete catalogue of genetic variation can identify signals previously missed and markedly increase the number of identified candidate functional alleles at each locus. Project data have been used to impute over 6 million genetic variants into GWAS, for traits as diverse as smoking[44] and multiple sclerosis[45], as an exclusionary filter in Mendelian disease studies[46] and tumour sequencing studies, and to design the next generation of genotyping arrays.

The results from this study also provide a template for future genome-wide sequencing studies on larger sample sets. Our plans for achieving the 1000 Genomes Project goals are described in Box 2. Other studies using phenotyped samples are already using components of the design and analysis framework described above.

Measurement of human DNA variation is an essential prerequisite for carrying out human genetics research. The 1000 Genomes Project represents a step towards a complete description of human DNA polymorphism. The larger data set provided by the full 1000 Genomes Project will allow more accurate imputation of variants in GWAS and thus better localization of disease-associated variants. The project will provide a template for studies using genome-wide sequence data. Applications of these data, and the methods developed to generate them, will contribute to a much more comprehensive understanding of the role of inherited DNA variation in human history, evolution and disease.

METHODS SUMMARY

The Supplementary Information provides full details of samples, data generation protocols, read mapping, SNP calling, short insertion and deletion calling, structural variation calling and *de novo* assembly. Details of methods used in the analyses relating to imputation, mutation rate estimation, functional annotation, population genetics and extrapolation to the full project are also presented.

1. The International Human Genome Sequencing Consortium. Finishing the euchromatic sequence of the human genome. *Nature* **431**, 931–945 (2004).
2. Sachidanandam, R. *et al.* A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. *Nature* **409**, 928–933 (2001).
3. The International HapMap Consortium. A haplotype map of the human genome. *Nature* **437**, 1299–1320 (2005).
4. The International HapMap Consortium. A second generation human haplotype map of over 3.1 million SNPs. *Nature* **449**, 851–861 (2007).
5. Hindorff, L. A., Junkins, H. A., Hall, P. N., Mehta, J. P. & Manolio, T. A. A catalog of published genome-wide association studies. 〈http://www.genome.gov/gwastudies〉 (2010).

6. Craddock, N. et al. Genome-wide association study of CNVs in 16,000 cases of eight common diseases and 3,000 shared controls. Nature 464, 713–720 (2010).

7. Manolio, T. A. et al. Finding the missing heritability of complex diseases. Nature 461, 747–753 (2009).

8. Nejentsev, S., Walker, N., Riches, D., Egholm, M. & Todd, J. A. Rare variants of IFIH1, a gene implicated in antiviral responses, protect against type 1 diabetes. Science 324, 387–389 (2009).

9. Cohen, J. C., Boerwinkle, E., Mosley, T. H. Jr & Hobbs, H. H. Sequence variations in PCSK9, low LDL, and protection against coronary heart disease. N. Engl. J. Med. 354, 1264–1272 (2006).

10. Levy, S. et al. The diploid genome sequence of an individual human. PLoS Biol. 5, e254 (2007).

11. Wheeler, D. A. et al. The complete genome of an individual by massively parallel DNA sequencing. Nature 452, 872–876 (2008).

12. Bentley, D. R. et al. Accurate whole human genome sequencing using reversible terminator chemistry. Nature 456, 53–59 (2008).

13. Wang, J. et al. The diploid genome sequence of an Asian individual. Nature 456, 60–65 (2008).

14. Li, H. et al. The sequence alignment/map format and SAMtools. Bioinformatics 25, 2078–2079 (2009).

15. Albers, C. et al. Dindel: Accurate indel calls from short read data. Genome Res. (in the press).

16. Lam, H. Y. et al. Nucleotide-resolution analysis of structural variants using BreakSeq and a breakpoint library. Nature Biotechnol. 28, 47–55 (2010).

17. The International HapMap 3 Consortium. Integrating common and rare genetic variation in diverse human populations. Nature 467, 52–58 (2010).

18. Conrad, D. F. et al. Origins and functional impact of copy number variation in the human genome. Nature 464, 704–712 (2010).

19. Irwin, J. A. et al. Investigation of heteroplasmy in the human mitochondrial DNA control region: a synthesis of observations from more than 5000 global population samples. J. Mol. Evol. 68, 516–527 (2009).

20. Balaresque, P. et al. A predominantly neolithic origin for European paternal lineages. PLoS Biol. 8, e1000285 (2010).

21. Wendl, M. C. & Wilson, R. K. The theory of discovering rare variants via DNA sequencing. BMC Genomics 10, 485 (2009).

22. Le, S. Q., Li, H. & Durbin, R. QCALL: SNP detection and genotyping from low coverage sequence data on multiple diploid samples. Genome Res. (in the press).

23. NHLBI Program for Genomic Applications. SeattleSNPs. ⟨http://pga.gs.washington.edu/⟩ (2010).

24. Xing, J. et al. Mobile elements create structural variation: analysis of a complete human genome. Genome Res. 19, 1516–1526 (2009).

25. Stranger, B. E. et al. Population genomics of human gene expression. Nature Genet. 39, 1217–1224 (2007).

26. Li, Y., Willer, C. J., Ding, J., Scheet, P. & Abecasis, G. R. MaCH: Using sequence and genotype data to estimate haplotypes and unobserved genotypes. Genet. Epi. (in the press).

27. Marchini, J. & Howie, B. Genotype imputation for genome-wide association studies. Nature Rev. Genet. 11, 499–511 (2010).

28. Dixon, A. L. et al. A genome-wide association study of global gene expression. Nature Genet. 39, 1202–1207 (2007).

29. Genovese, G. et al. Association of trypanolytic ApoL1 variants with kidney disease in African Americans. Science 329, 841–845 (2010).

30. Nachman, M. W. & Crowell, S. L. Estimate of the mutation rate per nucleotide in humans. Genetics 156, 297–304 (2000).

31. Kondrashov, A. S. Direct estimates of human per nucleotide mutation rates at 20 loci causing Mendelian diseases. Hum. Mutat. 21, 12–27 (2003).

32. Roach, J. C. et al. Analysis of genetic inheritance in a family quartet by whole-genome sequencing. Science 328, 636–639 (2010).

33. Charlesworth, B., Morgan, M. T. & Charlesworth, D. The effect of deleterious mutations on neutral molecular variation. Genetics 134, 1289–1303 (1993).

34. Maynard Smith, J. & Haigh, J. The hitch-hiking effect of a favourable gene. Genet. Res. 23, 23–35 (1974).

35. Cai, J. J., Macpherson, J. M., Sella, G. & Petrov, D. A. Pervasive hitchhiking at coding and regulatory sites in humans. PLoS Genet. 5, e1000336 (2009).

36. Voight, B. F., Kudaravalli, S., Wen, X. & Pritchard, J. K. A map of recent positive selection in the human genome. PLoS Biol. 4, e72 (2006).

37. Barreiro, L. B., Laval, G., Quach, H., Patin, E. & Quintana-Murci, L. Natural selection has driven population differentiation in modern humans. Nature Genet. 40, 340–345 (2008).

38. Lamason, R. L. et al. SLC24A5, a putative cation exchanger, affects pigmentation in zebrafish and humans. Science 310, 1782–1786 (2005).

39. Tournamille, C., Colin, Y., Cartron, J. P. & Le Van Kim, C. Disruption of a GATA motif in the Duffy gene promoter abolishes erythroid gene expression in Duffy-negative individuals. Nature Genet. 10, 224–228 (1995).

40. Myers, S. et al. Drive against hotspot motifs in primates implicates the PRDM9 gene in meiotic recombination. Science 327, 876–879 (2010).

41. Myers, S., Freeman, C., Auton, A., Donnelly, P. & McVean, G. A common sequence motif associated with recombination hot spots and genome instability in humans. Nature Genet. 40, 1124–1129 (2008).

42. Baudat, F. et al. PRDM9 is a major determinant of meiotic recombination hotspots in humans and mice. Science 327, 836–840 (2010).

43. Parvanov, E. D., Petkov, P. M. & Paigen, K. Prdm9 controls activation of mammalian recombination hotspots. Science 327, 835 (2010).

44. Liu, J. Z. et al. Meta-analysis and imputation refines the association of 15q25 with smoking quantity. Nature Genet. 42, 436–440 (2010).

45. Sanna, S. et al. Variants within the immunoregulatory CBLB gene are associated with multiple sclerosis. Nature Genet. 42, 495–497 (2010).

46. Musunuru, K. et al. Exome sequencing, mutations in ANGPTL3, and familial combined hypolipidemia. N. Engl. J. Med. (in the press).

47. Ewing, A. D. & Kazazian, H. H. Jr. High-throughput sequencing reveals extensive variation in human-specific L1 content in individual human genomes. Genome Res. 20, 1262–1270 (2010).

48. Mills, R. E. et al. An initial map of insertion and deletion (INDEL) variation in the human genome. Genome Res. 16, 1182–1190 (2006).

49. Liti, G. et al. Population genomics of domestic and wild yeasts. Nature 458, 337–341 (2009).

50. Li, Y., Willer, C., Sanna, S. & Abecasis, G. Genotype imputation. Annu. Rev. Genomics Hum. Genet. 10, 387–406 (2009).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Author Contributions Details of author contributions can be found in the author list.

The 1000 Genomes Consortium (Participants are arranged by project role, then by institution alphabetically, and finally alphabetically within institutions except for Principal Investigators and Project Leaders, as indicated.)

Corresponding author Richard M. Durbin[1]

Steering committee David L. Altshuler[2,3,4] (Co-Chair), Richard M. Durbin[1] (Co-Chair), Gonçalo R. Abecasis[5], David R. Bentley[6], Aravinda Chakravarti[7], Andrew G. Clark[8], Francis S. Collins[9], Francisco M. De La Vega[10], Peter Donnelly[11], Michael Egholm[12], Paul Flicek[13], Stacey B. Gabriel[2], Richard A. Gibbs[14], Bartha M. Knoppers[15], Eric S. Lander[2], Hans Lehrach[16], Elaine R. Mardis[17], Gil A. McVean[11,18], Debbie A. Nickerson[19],

Leena Peltonen‡, Alan J. Schafer[20], Stephen T. Sherry[21], Jun Wang[22,23], Richard K. Wilson[17]

Production group: Baylor College of Medicine Richard A. Gibbs[14] (Principal Investigator), David Deiros[14], Mike Metzker[14], Donna Muzny[14], Jeff Reid[14], David Wheeler[14]; **BGI-Shenzhen** Jun Wang[22,23] (Principal Investigator), Jingxiang Li[22], Min Jian[22], Guoqing Li[22], Ruiqiang Li[22,23], Huiqing Liang[22], Geng Tian[22], Bo Wang[22], Jian Wang[22], Wei Wang[22], Huanming Yang[22], Xiuqing Zhang[22], Huisong Zheng[22]; **Broad Institute of MIT and Harvard** Eric S. Lander[2] (Principal Investigator), David L. Altshuler[2,3,4], Lauren Ambrogio[2], Toby Bloom[2], Kristian Cibulskis[2], Tim J. Fennell[2], Stacey B. Gabriel[2] (Co-Chair), David B. Jaffe[2], Erica Shefler[2], Carrie L. Sougnez[2]; **Illumina** David R. Bentley[6] (Principal Investigator), Niall Gormley[6], Sean Humphray[6], Zoya Kingsbury[6], Paula Koko-Gonzales[6], Jennifer Stone[6]; **Life Technologies** Kevin J. McKernan[24] (Principal Investigator), Gina L. Costa[24], Jeffry K. Ichikawa[24], Clarence C. Lee[24]; **Max Planck Institute for Molecular Genetics** Ralf Sudbrak[16] (Project Leader), Hans Lehrach[16] (Principal Investigator), Tatiana A. Borodina[16], Andreas Dahl[25], Alexey N. Davydov[16], Peter Marquardt[16], Florian Mertes[16], Wilfried Nietfeld[16], Philip Rosenstiel[26], Stefan Schreiber[26], Aleksey V. Soldatov[16], Bernd Timmermann[16], Marius Tolzmann[16]; **Roche Applied Science** Michael Egholm[12] (Principal Investigator), Jason Affourtit[27], Dana Ashworth[27], Said Attiya[27], Melissa Bachorski[27], Eli Buglione[27], Adam Burke[27], Amanda Caprio[27], Christopher Celone[27], Shauna Clark[27], David Conners[27], Brian Desany[27], Lisa Gu[27], Lorri Guccione[27], Kalvin Kao[27], Andrew Kebbel[27], Jennifer Knowlton[27], Matthew Labrecque[27], Louise McDade[27], Craig Mealmaker[27], Melissa Minderman[27], Anne Nawrocki[27], Faheem Niazi[27], Kristen Pareja[27], Ravi Ramenani[27], David Riches[27], Wanmin Song[27], Cynthia Turcotte[27], Shally Wang[27]; **Washington University in St Louis** Elaine R. Mardis[17] (Co-Chair) (Co-Principal Investigator), Richard K. Wilson[17] (Co-Principal Investigator), David Dooling[17], Lucinda Fulton[17], Robert Fulton[17], George Weinstock[17]; **Wellcome Trust Sanger Institute** Richard M. Durbin[1] (Principal Investigator), John Burton[1], David M. Carter[1], Carol Churcher[1], Alison Coffey[1], Anthony Cox[1], Aarno Palotie[1,28], Michael Quail[1], Tom Skelly[1], James Stalker[1], Harold P. Swerdlow[1], Daniel Turner[1]

Analysis group: Agilent Technologies Anniek De Witte[29], Shane Giles[29]; **Baylor College of Medicine** Richard A. Gibbs[14] (Principal Investigator), David Wheeler[14], Matthew Bainbridge[14], Danny Challis[14], Aniko Sabo[14], Fuli Yu[14], Jin Yu[14]; **BGI-Shenzhen** Jun Wang[22,23] (Principal Investigator), Xiaodong Fang[22], Xiaosen Guo[22], Ruiqiang Li[22,23], Yingrui Li[22], Ruibang Luo[22], Shuaishuai Tai[22], Honglong Wu[22], Hancheng Zheng[22], Xiaole Zheng[22], Yan Zhou[22], Guoqing Li[22], Jian Wang[22], Huanming Yang[22]; **Boston College** Gabor T. Marth[30] (Principal Investigator), Erik P. Garrison[30], Weichun Huang[31], Amit Indap[30], Deniz Kural[30], Wan-Ping Lee[30], Wen Fung Leong[30], Aaron R. Quinlan[32], Chip Stewart[30], Michael P. Stromberg[33], Alistair N. Ward[30], Jiantao Wu[30]; **Brigham and Women's Hospital** Charles Lee[34] (Principal Investigator), Ryan E. Mills[34], Xinghua Shi[34]; **Broad Institute of MIT and Harvard** Mark J. Daly[2] (Principal Investigator), Mark A. DePristo[2] (Project Leader), David L. Altshuler[2,34], Aaron D. Ball[2], Eric Banks[2], Toby Bloom[2], Brian L. Browning[35], Kristian Cibulskis[2], Tim J. Fennell[2], Kiran V. Garimella[2], Sharon R. Grossman[2,36], Robert E. Handsaker[2], Matt Hanna[2], Chris Hartl[2], David B. Jaffe[2], Andrew M. Kernytsky[2], Joshua M. Korn[2], Heng Li[2], Jared R. Maguire[2], Steven A. McCarroll[2,4], Aaron McKenna[2], James C. Nemesh[2], Anthony A. Philippakis[2], Ryan E. Poplin[2], Alkes Price[37], Manuel A. Rivas[2], Pardis C. Sabeti[2,36], Stephen F. Schaffner[2], Erica Shefler[2], Ilya A. Shlyakhter[2,36]; **Cardiff University, The Human Gene Mutation Database** David N. Cooper[38] (Principal Investigator), Edward V. Ball[38], Matthew Mort[38], Andrew D. Phillips[38], Peter D. Stenson[38]; **Cold Spring Harbor Laboratory** Jonathan Sebat[39] (Principal Investigator), Vladimir Makarov[40], Kenny Ye[41], Seungtai C. Yoon[42]; **Cornell and Stanford Universities** Carlos D. Bustamante[43] (Co-Principal Investigator), Andrew G. Clark[8] (Co-Principal Investigator), Adam Boyko[43], Jeremiah Degenhardt[8], Simon Gravel[43], Ryan N. Gutenkunst[44], Mark Kaganovich[43], Alon Keinan[8], Phil Lacroute[43], Xin Ma[8], Andy Reynolds[8]; **European Bioinformatics Institute** Laura Clarke[13] (Project Leader), Paul Flicek[13] (Co-Chair, DCC) (Principal Investigator), Fiona Cunningham[13], Javier Herrero[13], Stephen Keenen[13], Eugene Kulesha[13], Rasko Leinonen[13], William M. McLaren[13], Rajesh Radhakrishnan[13], Richard Smith[13], Vadim Zalunin[13], Xiangqun Zheng-Bradley[13]; **European Molecular Biology Laboratory** Jan O. Korbel[45] (Principal Investigator), Adrian M. Stütz[45]; **Illumina** Sean Humphray[6] (Project Leader), Markus Bauer[6], R. Keira Cheetham[6], Tony Cox[6], Michael Eberle[6], Terena James[6], Scott Kahn[6], Lisa Murray[6]; **Johns Hopkins University** Aravinda Chakravarti[7]; **Leiden University Medical Center** Kai Ye[46]; **Life Technologies** Francisco M. De La Vega[10] (Principal Investigator), Yutao Fu[24], Fiona C. L. Hyland[10], Jonathan M. Manning[24], Stephen F. McLaughlin[24], Heather E. Peckham[24], Onur Sakarya[10], Yongming A. Sun[10], Eric F. Tsung[24]; **Louisiana State University** Mark A. Batzer[47] (Principal Investigator), Miriam K. Konkel[47], Jerilyn A. Walker[47]; **Max Planck Institute for Molecular Genetics** Ralf Sudbrak[16] (Project Leader), Marcus H. Albrecht[16], Vyacheslav S. Amstislavskiy[16], Ralf Herwig[16], Dimitri V. Parkhomchuk[16]; **US National Institutes of Health** Stephen T. Sherry[21] (Co-Chair, DCC) (Principal Investigator), Richa Agarwala[21], Hoda M. Khouri[21], Aleksandr O. Morgulis[21], Justin E. Paschall[21], Lon D. Phan[21], Kirill E. Rotmistrovsky[21], Robert D. Sanders[21], Martin F. Shumway[21], Chunlin Xiao[21]; **Oxford University** Gil A. McVean[11,18] (Co-Chair) (Co-Chair, Population Genetics) (Principal Investigator), Adam Auton[11], Zamin Iqbal[11], Gerton Lunter[11], Jonathan L. Marchini[11,18], Loukas Moutsianas[18], Simon Myers[11,18], Afidalina Tumian[18]; **Roche Applied Science** Brian Desany[27] (Project Leader), James Knight[27], Roger Winer[27]; **The Translational Genomics Research Institute** David W. Craig[48] (Principal Investigator), Steve M. Beckstrom-Sternberg[48], Alexis Christoforides[48], Ahmet A. Kurdoglu[48], John V. Pearson[48], Shripad A. Sinari[48], Waibhav D. Tembe[48]; **University of California, Santa Cruz** David Haussler[49] (Principal Investigator), Angie S. Hinrichs[49], Sol J. Katzman[49], Andrew Kern[49], Robert M. Kuhn[49]; **University of Chicago** Molly Przeworski[50] (Co-Chair, Population Genetics) (Principal Investigator), Ryan D. Hernandez[51], Bryan Howie[52], Joanna L. Kelley[52], S. Cord Melton[52]; **University of Michigan** Gonçalo R. Abecasis[5] (Co-Chair) (Principal Investigator), Yun Li[5] (Project Leader), Paul Anderson[5], Tom Blackwell[5], Wei Chen[5], William O. Cookson[53], Jun Ding[5], Hyun Min Kang[5], Mark Lathrop[54], Liming Liang[55], Miriam F. Moffatt[53], Paul Scheet[56], Carlo Sidore[5], Matthew Snyder[5], Xiaowei Zhan[5], Sebastian Zöllner[5]; **University of Montreal** Philip Awadalla[57] (Principal Investigator), Ferran Casals[58], Youssef Idaghdour[58], John Keebler[58], Eric A. Stone[58], Martine Zilversmit[58]; **University of Utah** Lynn Jorde[59] (Principal Investigator), Jinchuan Xing[59]; **University of Washington** Evan E. Eichler[60] (Principal Investigator), Can Alkan[60], Iman Hajirasouliha[61], Fereydoun Hormozdiari[61], Jeffrey M. Kidd[19,43], S. Cenk Sahinalp[61], Peter H. Sudmant[19]; **Washington University in St Louis** Elaine R. Mardis[17] (Co-Principal Investigator), Ken Chen[17], Asif Chinwalla[17], Li Ding[17], Daniel C. Koboldt[17], Mike D. McLellan[17], David Dooling[17], George Weinstock[17], John W. Wallis[17], Michael C. Wendl[17], Qunyuan Zhang[17]; **Wellcome Trust Sanger Institute** Richard M. Durbin[1] (Principal Investigator), Cornelis A. Albers[62], Qasim Ayub[1], Senduran Balasubramaniam[1], Jeffrey C. Barrett[1], David M. Carter[1], Yuan Chen[1], Donald F. Conrad[1], Petr Danecek[1], Emmanouil T. Dermitzakis[63], Min Hu[1], Ni Huang[1], Matt E. Hurles[1], Hanjun Jin[64], Luke Jostins[1], Thomas M. Keane[1], Si Quang Le[1], Sarah Lindsay[1], Quan Long[1], Daniel G. MacArthur[1], Stephen B. Montgomery[63], Leopold Parts[1], James Stalker[1], Chris Tyler-Smith[1], Klaudia Walter[1], Yujun Zhang[1]; **Yale and Stanford Universities** Mark B. Gerstein[65,66] (Co-Principal Investigator), Michael Snyder[43] (Co-Principal Investigator), Alexej Abyzov[65], Suganthi Balasubramanian[67], Robert Bjornson[66], Jiang Du[66], Fabian Grubert[43], Lukas Habegger[65], Rajini Haraksingh[65], Justin Jee[65], Ekta Khurana[67], Hugo Y. K. Lam[43], Jing Leng[65], Xinmeng Jasmine Mu[65], Alexander E. Urban[43,68], Zhengdong Zhang[67]

Structural variation group: BGI-Shenzhen Yingrui Li[22], Ruibang Luo[22]; **Boston College** Gabor T. Marth[30] (Principal Investigator), Erik P. Garrison[30], Deniz Kural[30], Aaron R. Quinlan[32], Chip Stewart[30], Michael P. Stromberg[33], Alistair N. Ward[30], Jiantao Wu[30]; **Brigham and Women's Hospital** Charles Lee[34] (Co-Chair) (Principal Investigator), Ryan E. Mills[34], Xinghua Shi[34]; **Broad Institute of MIT and Harvard** Steven A. McCarroll[2,4] (Project Leader), Mark A. DePristo[2], Robert E. Handsaker[2], Chris Hartl[2], Joshua M. Korn[2], Heng Li[2], James C. Nemesh[2]; **Cold Spring Harbor Laboratory** Jonathan Sebat[39] (Principal Investigator), Vladimir Makarov[40], Kenny Ye[41], Seungtai C. Yoon[42]; **Cornell and Stanford Universities** Jeremiah Degenhardt[8], Mark Kaganovich[43]; **European Bioinformatics Institute** Laura Clarke[13] (Project Leader), Richard E. Smith[13], Xiangqun Zheng-Bradley[13]; **European Molecular Biology Laboratory** Jan O. Korbel[45] (Principal Investigator); **Illumina** Sean Humphray[6] (Project Leader), R. Keira Cheetham[6], Michael Eberle[6], Scott Kahn[6], Lisa Murray[6]; **Leiden University Medical Center** Kai Ye[46]; **Life Technologies** Francisco M. De La Vega[10] (Principal Investigator), Yutao Fu[24], Heather E. Peckham[24], Yongming A. Sun[10]; **Louisiana State University** Mark A. Batzer[47] (Principal Investigator), Miriam K. Konkel[47], Jerilyn A. Walker[47]; **US National Institutes of Health** Chunlin Xiao[21]; **Oxford University** Zamin Iqbal[11]; **Roche Applied Science** Brian Desany[27]; **University of Michigan** Tom Blackwell[5] (Project Leader), Matthew Snyder[5]; **University of Utah** Jinchuan Xing[59]; **University of Washington** Evan E. Eichler[60] (Co-Chair) (Principal Investigator), Gozde Aksay[19], Can Alkan[60], Iman Hajirasouliha[61], Fereydoun Hormozdiari[61], Jeffrey M. Kidd[19,43]; **Washington University in St Louis** Ken Chen[17], Asif Chinwalla[17], Li Ding[17], Mike D. McLellan[17], John W. Wallis[17]; **Wellcome Trust Sanger Institute** Matt E. Hurles[1] (Co-Chair) (Principal Investigator), Donald F. Conrad[1], Klaudia Walter[1], Yujun Zhang[1]; **Yale and Stanford Universities** Mark B. Gerstein[65,66] (Co-Principal Investigator), Michael Snyder[43] (Co-Principal Investigator), Alexej Abyzov[65], Jiang Du[66], Fabian Grubert[43], Rajini Haraksingh[65], Justin Jee[65], Ekta Khurana[67], Hugo Y. K. Lam[43], Jing Leng[65], Xinmeng Jasmine Mu[65], Alexander E. Urban[43,68], Zhengdong Zhang[67]

Exon pilot group: Baylor College of Medicine Richard A. Gibbs[14] (Co-Chair) (Principal Investigator), Matthew Bainbridge[14], Danny Challis[14], Cristian Coafra[14], Huyen Dinh[14], Christie Kovar[14], Sandy Lee[14], Donna Muzny[14], Lynne Nazareth[14], Jeff Reid[14], Aniko Sabo[14], Fuli Yu[14], Jin Yu[14]; **Boston College** Gabor T. Marth[30] (Co-Chair) (Principal Investigator), Erik P. Garrison[30], Amit Indap[30], Wen Fung Leong[30], Aaron R. Quinlan[32], Chip Stewart[30], Alistair N. Ward[30], Jiantao Wu[30]; **Broad Institute of MIT and Harvard** Kristian Cibulskis[2], Tim J. Fennell[2], Stacey B. Gabriel[2], Kiran V. Garimella[2], Chris Hartl[2], Erica Shefler[2], Carrie L. Sougnez[2], Jane Wilkinson[2]; **Cornell and Stanford Universities** Andrew G. Clark[8] (Co-Principal Investigator), Simon Gravel[43], Fabian Grubert[43]; **European Bioinformatics Institute** Laura Clarke[13] (Project Leader), Paul Flicek[13] (Principal Investigator), Richard E. Smith[13], Xiangqun Zheng-Bradley[13]; **US National Institutes of Health** Stephen T. Sherry[21] (Principal Investigator), Hoda M. Khouri[21], Justin E. Paschall[21], Martin F. Shumway[21], Chunlin Xiao[21]; **Oxford University** Gil A. McVean[11,18]; **University of California, Santa Cruz** Sol J. Katzman[49]; **University of Michigan** Gonçalo R. Abecasis[5] (Principal Investigator), Tom Blackwell[5]; **Washington University in St Louis** Elaine R. Mardis[17] (Principal Investigator), David Dooling[17], Lucinda Fulton[17], Robert Fulton[17], Daniel C. Koboldt[17]; **Wellcome Trust Sanger Institute** Richard M. Durbin[1] (Principal Investigator), Senduran Balasubramaniam[1], Allison Coffey[1], Thomas M. Keane[1], Daniel G. MacArthur[1], Aarno Palotie[1,28], Carol Scott[1], James Stalker[1], Chris Tyler-Smith[1]; **Yale University** Mark B. Gerstein[65,66] (Principal Investigator), Suganthi Balasubramanian[67]

Samples and ELSI group Aravinda Chakravarti[7] (Co-Chair), Bartha M. Knoppers[15] (Co-Chair), Leena Peltonen‡ (Co-Chair), Gonçalo R. Abecasis[5], Carlos D. Bustamante[43], Neda Gharani[69], Richard A. Gibbs[14], Lynn Jorde[59], Jane S. Kaye[70], Alastair Kent[71], Taosha Li[22], Amy L. McGuire[72], Gil A. McVean[11,18], Pilar N. Ossorio[73], Charles N. Rotimi[74], Yeyang Su[22], Lorraine H. Toji[69], Chris Tyler-Smith[1]

Scientific management Lisa D. Brooks[75], Adam L. Felsenfeld[75], Jean E. McEwen[75], Assya Abdallah[76], Christopher R. Juenger[77], Nicholas C. Clemm[75], Francis S. Collins[9], Audrey Duncanson[20], Eric D. Green[78], Mark S. Guyer[75], Jane L. Peterson[75], Alan J. Schafer[20]

Writing group Gonçalo R. Abecasis[5], David L. Altshuler[2,3,4], Adam Auton[11], Lisa D. Brooks[75], Richard M. Durbin[1], Richard A. Gibbs[14], Matt E. Hurles[1], Gil A. McVean[11,18]

[1]Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Cambridge CB10 1SA, UK. [2]The Broad Institute of MIT and Harvard, 7 Cambridge Center, Cambridge, Massachusetts 02142, USA. [3]Center for Human Genetic Research, Massachusetts General Hospital, Boston, Massachusetts 02114, USA. [4]Department of Genetics, Harvard Medical School, Cambridge, Massachusetts 02115, USA. [5]Center for Statistical Genetics and Biostatistics, University of Michigan, Ann Arbor, Michigan 48109, USA. [6]Illumina Cambridge Ltd, Chesterford Research Park, Little Chesterford, Nr Saffron Walden, Essex CB10 1XL, UK. [7]McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University School of Medicine, Baltimore, Maryland 21205, USA. [8]Center for Comparative and Population Genomics, Cornell University, Ithaca, New York 14850, USA. [9]US National Institutes of Health, 1 Center Drive, Bethesda, Maryland 20892, USA. [10]Life Technologies, Foster City, California 94404, USA. [11]Wellcome Trust Centre for Human Genetics, Roosevelt Drive, Oxford OX3 7BN, UK. [12]Pall Corporation, 25 Harbor Park Drive, Port Washington, New York 11050, USA. [13]European Bioinformatics Institute, Wellcome Trust Genome Campus, Cambridge CB10 1SD, UK. [14]Human Genome Sequencing Center, Baylor College of Medicine, 1 Baylor Plaza, Houston, Texas 77030, USA. [15]Centre of Genomics and Policy, McGill University, Montréal, Québec H3A 1A4, Canada. [16]Max Planck Institute for Molecular Genetics, D-14195 Berlin-Dahlem, Germany. [17]The Genome Center, Washington University School of Medicine, St Louis, Missouri 63108, USA. [18]Department of Statistics, University of Oxford, Oxford OX1 3TG, UK. [19]Department of Genome Sciences, University of Washington School of Medicine, Seattle, Washington 98195, USA. [20]Wellcome Trust, Gibbs Building, 215 Euston Road, London NW1 2BE, UK. [21]US National Institutes of Health, National Center for Biotechnology Information, 45 Center Drive, Bethesda, Maryland 20892, USA. [22]BGI-Shenzhen, Shenzhen 518083, China. [23]Department of Biology, University of Copenhagen 2200, Denmark. [24]Life Technologies, Beverly, Massachusetts 01915, USA. [25]Deep Sequencing Group, Biotechnology Center TU Dresden, Tatzberg 47/49, 01307 Dresden, Germany. [26]Institute of Clinical Molecular Biology, Christian-Albrechts-University Kiel, Kiel 24105, Germany. [27]Roche Applied Science, 20 Commercial Street, Branford, Connecticut 06405, USA. [28]Department of Medical Genetics, Institute of Molecular Medicine (FIMM) of the University of Helsinki and Helsinki University Hospital, Helsinki 00290, Finland. [29]Agilent Technologies Inc., Santa Clara, California 95051, USA. [30]Department of Biology, Boston College, Chestnut Hill, Massachusetts 02467, USA. [31]US National Institutes of Health, National Institute of Environmental Health Sciences, 111 T W Alexander Drive, Research Triangle Park, North Carolina 27709, USA. [32]Department of Biochemistry and Molecular Genetics, University of Virginia School of Medicine, Charlottesville, Virginia 22908, USA. [33]Illumina, San Diego, California 92121, USA. [34]Department of Pathology, Brigham and Women's Hospital and Harvard Medical School, Boston, Massachusetts 02115, USA. [35]Department of Medicine, Division of Medical Genetics, University of Washington, Seattle, Washington 98195, USA. [36]Center for Systems Biology, Department of Organismic and Evolutionary Biology, Harvard University, Cambridge, Massachusetts 02138, USA. [37]Department of Epidemiology, Harvard School of Public Health, Boston, Massachusetts 02115, USA. [38]Institute of Medical Genetics, Cardiff University, Heath Park, Cardiff CF14 4XN, UK. [39]Departments of Psychiatry and Cellular and Molecular Medicine, University of California San Diego, 9500 Gilman Drive, La Jolla, California 92093, USA. [40]Seaver Autism Center and Department of Psychiatry, Mount Sinai School of Medicine, New York, New York 10029, USA. [41]Department of Epidemiology and Population Health, Albert Einstein College of Medicine, Bronx, New York 10461, USA. [42]Department of Genetics and Genomic Sciences, Mount Sinai School of Medicine, New York, New York 10029, USA. [43]Department of Genetics, Stanford University, Stanford, California 94305, USA. [44]Department of Molecular and Cellular Biology, University of Arizona, Tucson, Arizona 85721, USA. [45]European Molecular Biology Laboratory, Genome Biology Research Unit, Meyerhofstrasse 1, Heidelberg 69117, Germany. [46]Molecular Epidemiology Section, Medical Statistics and Bioinformatics, Leiden University Medical Center, 2333 ZA, The Netherlands. [47]Department of Biological Sciences, Louisiana State University, Baton Rouge, Louisiana 70803, USA. [48]The Translational Genomics Research Institute, 445 N Fifth Street, Phoenix, Arizona 85004, USA. [49]Center for Biomolecular Science and Engineering, University of California Santa Cruz, Santa Cruz, California 95064, USA. [50]Department of Human Genetics and Howard Hughes Medical Institute, University of Chicago, Chicago, Illinois 60637, USA. [51]Department of Bioengineering and Therapeutic Sciences, University of California San Francisco, San Francisco, California 94158, USA. [52]Department of Human Genetics, University of Chicago, Chicago, Illinois 60637, USA. [53]National Heart and Lung Institute, Imperial College London, London SW7 2, UK. [54]Centre Nationale de Génotypage, Evry 91000, France. [55]Departments of Epidemiology and Biostatistics, Harvard School of Public Health, Boston, Massachusetts 02115, USA. [56]Department of Epidemiology, University of Texas MD Anderson Cancer Center, Houston, Texas 77030, USA. [57]Department of Pediatrics, Faculty of Medicine, University of Montréal, Ste. Justine Hospital Research Centre, Montréal, Québec H3T 1C5, Canada. [58]Department of Medicine, Centre Hospitalier de l'Université de Montréal Research Center, Université de Montréal, Montréal, Québec H2L 2W5, Canada. [59]Eccles Institute of Human Genetics, University of Utah School of Medicine, Salt Lake City, Utah 84112, USA. [60]Department of Genome Sciences, University of Washington School of Medicine and Howard Hughes Medical Institute, Seattle, Washington 98195, USA. [61]Department of Computer Science, Simon Fraser University, Burnaby, British Columbia V5A 1S6, Canada. [62]Department of Haematology, University of Cambridge and National Health Service Blood and Transplant, Cambridge CB2 1TN, UK. [63]Department of Genetic Medicine and Development, University of Geneva Medical School, Geneva 1211, Switzerland. [64]Center for Genome Science, Korea National Institute of Health, 194, Tongil-Lo, Eunpyung-Gu, Seoul 122-701, Korea. [65]Program in Computational Biology and Bioinformatics, Yale University, New Haven, Connecticut 06520, USA. [66]Department of Computer Science, Yale University, New Haven, Connecticut 06520, USA. [67]Department of Molecular Biophysics and Biochemistry, Yale University, New Haven, Connecticut 06520, USA. [68]Department of Psychiatry and Behavioral Studies, Stanford University, Stanford, California 94305, USA. [69]Coriell Institute, 403 Haddon Avenue, Camden, New Jersey 08103, USA. [70]Centre for Health, Law and Emerging Technologies, University of Oxford, Oxford OX3 7LF, UK. [71]Genetic Alliance, 436 Essex Road, London N1 3QP, UK. [72]Center for Medical Ethics and Health Policy, Baylor College of Medicine, 1 Baylor Plaza, Houston, Texas 77030, USA. [73]Department of Medical History and Bioethics, University of Wisconsin–Madison, Madison, Wisconsin 53706, USA. [74]US National Institutes of Health, Center for Research on Genomics and Global Health, 12 South Drive, Bethesda, Maryland 20892, USA. [75]US National Institutes of Health, National Human Genome Research Institute, 5635 Fishers Lane, Bethesda, Maryland 20892, USA. [76]The George Washington University School of Medicine and Health Sciences, Washington DC 20037, USA. [77]US Food and Drug Administration, 11400 Rockville Pike, Rockville, Maryland 20857, USA. [78]US National Institutes of Health, National Human Genome Research Institute, 31 Center Drive, Bethesda, Maryland 20892, USA. ‡Deceased.

ARTICLE

Homologue structure of the SLAC1 anion channel for closing stomata in leaves

Yu-hang Chen[1,2], Lei Hu[3], Marco Punta[2,4], Renato Bruni[2], Brandan Hillerich[2], Brian Kloss[2], Burkhard Rost[1,2,4], James Love[2], Steven A. Siegelbaum[3,5,6] & Wayne A. Hendrickson[1,2,6,7]

The plant SLAC1 anion channel controls turgor pressure in the aperture-defining guard cells of plant stomata, thereby regulating the exchange of water vapour and photosynthetic gases in response to environmental signals such as drought or high levels of carbon dioxide. Here we determine the crystal structure of a bacterial homologue (*Haemophilus influenzae*) of SLAC1 at 1.20 Å resolution, and use structure-inspired mutagenesis to analyse the conductance properties of SLAC1 channels. SLAC1 is a symmetrical trimer composed from quasi-symmetrical subunits, each having ten transmembrane helices arranged from helical hairpin pairs to form a central five-helix transmembrane pore that is gated by an extremely conserved phenylalanine residue. Conformational features indicate a mechanism for control of gating by kinase activation, and electrostatic features of the pore coupled with electrophysiological characteristics indicate that selectivity among different anions is largely a function of the energetic cost of ion dehydration.

Stomatal pores in the leaves of plants permit the influx of atmospheric CO_2 in exchange for transpirational evaporation of water[1,2]. A pair of kidney-shaped guard cells defines each pore aperture, and turgor pressure variation in these cells determines the degree of stomatal pore openness. Depending on diverse environmental factors, the stomata close to prevent H_2O loss and open to admit CO_2 for photosynthesis. Environmental stimuli that lead to stomatal closure include darkness, high CO_2 levels, ozone, low air humidity and drought. The plant hormone abscisic acid (ABA) is critical for signal transduction from these stimuli. Mutational screens in *Arabidopsis thaliana* for CO_2 and ozone sensitivity identified a protein with ten predicted transmembrane helices, now called SLOW ANION CHANNEL 1 (SLAC1), as having a central role in the control of stomatal closure[3–5]. Recent studies proved that SLAC1 is indeed an anion channel[6,7], with characteristics like those of slow anion channels found in guard cells[8], and that it is activated by phosphorylation from the OST1 kinase[9]. OST1 activity is negatively regulated by the ABI1 phosphatase[10,11], which is in turn inhibited by the stomatal ABA receptors PYR and RCAR[12,13] when in the ternary hormone–receptor–phosphatase complex[14–18]. Thereby, ABA stimulates SLAC1 channel activity. Resulting Cl^- efflux through SLAC1 causes membrane depolarization, which activates outward-rectifying K^+ channels, leading to KCl and water efflux to reduce turgor further and cause stomatal closure.

SLAC1 expression in *Arabidopsis* is confined to the guard cells of leaves, but other *Arabidopsis* tissues do have SLAC1 homologues[3], named SLAH1–SLAH4. The identifying mutations *slac1-1* (ref. 4) and *slac1-2* (ref. 3) are, respectively, in predicted transmembrane segments 9 (S456F) and 1 (G194D) of a protein that includes substantial amino- and carboxy-terminal extensions outside a 10-helix transmembrane domain. SLAH1, which is absent from leaves and lacks the terminal extensions of SLAC1, fully complements the mutant phenotype in *slac1-2* guard cell protoplasts[3]. *SLAC1* and homologues are also present in other plant genomes, including nine in rice (*Oryza sativa*) and five in grapevines (*Vitis vinifera*). SLAC1
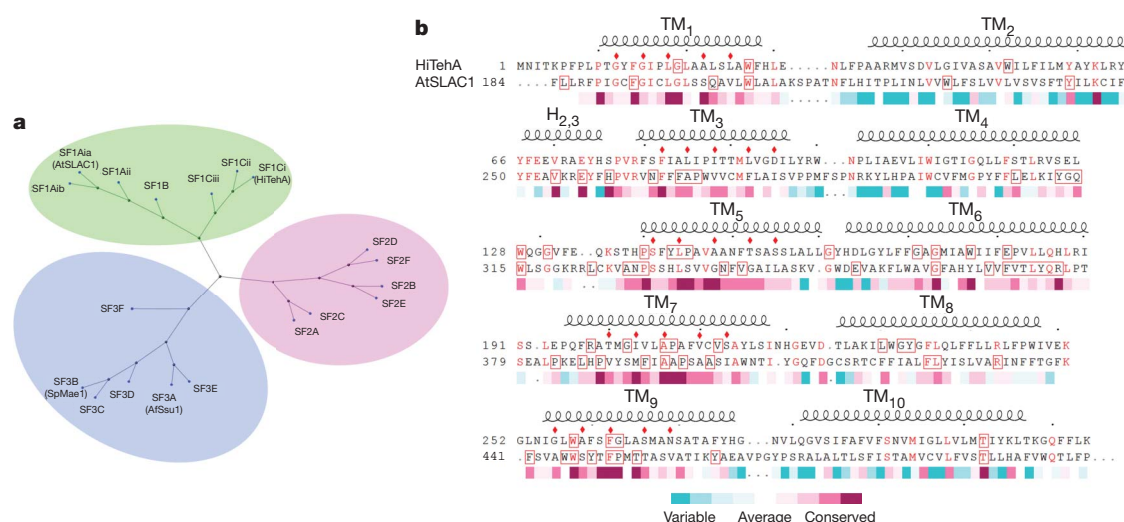
relatives, some quite remote, also occur in bacteria, archaea and fungi. Known prokaryotic homologues contain only the predicted transmembrane domain of SLAC1, but some fungal homologues do have N- and C-terminal extensions. One homologue, Mae1 from the yeast *Schizosaccharomyces pombe*, functions as a malate uptake transporter[19]; another, Ssu1 from *Saccharomyces cerevisiae* and other fungi including *Aspergillus fumigatus*, is characterized as a sulphite efflux pump[20,21]; and TehA from *Escherichia coli* is identified as a tellurite resistance protein by virtue of its association in the *tehA/tehB* operon[22,23]. Despite a lack of further biochemical characterization, many homologues are annotated as tellurite resistance/dicarboxylate transporter (TDT) proteins.

We have undertaken structural and functional characterizations of the SLAC1 anion channel. We first solved an atomic-resolution crystal structure of the TehA homologue from *Haemophilus influenzae*, and we then developed a homology model for *Arabidopsis* SLAC1. This model allowed us to conduct mutagenesis for functional testing of structure-inspired hypotheses on gating and selectivity. We expressed *Haemophilus* TehA and *Arabidopsis* SLAC1 in *Xenopus* oocytes to characterize channel properties of these proteins and mutant variants. We also determined crystal structures for several mutant variants, including the homologue of *slac1-2*.

Structure of SLAC1 bacterial homologue TehA

We performed a bioinformatic analysis of SLAC1-related proteins, first clustering nearly 900 non-redundant sequences into a superfamily at the PSI-BLAST level $E \leq 10^{-3}$, then into three distinct families at an initial threshold of $E \leq 10^{-30}$, and finally into subfamilies at a typical initial threshold of $E \leq 10^{-55}$. Because previous annotation is not well founded in experiment and SLAC1 is now the best-characterized member, we adopt a nomenclature defining a SLAC superfamily divided into families identified as SF1–SF3 and subfamilies SF1A, SF1B, etc. Family SF1 comprises the plant SLAC proteins and close bacterial homologues; family SF2 comprises a distinct set of bacterial proteins often annotated

Figure 1 | Sequence analysis for the SLAC1 superfamily. a, Family tree. The presentation was computed by the program COBALT[46] from representative subfamily sequences (Supplementary Tables 1 and 2), including *Arabidopsis thaliana* SLAC1 for SF1Ai, *Haemophilus influenzae* TehA for SF1Ci, *Escherichia coli* TehA for SF1Cii, *Vibrio parahaemolyticus* for SF2A, *Staphylococcus aureus* for SF2B, *Aspergillus fumigatus* Ssu1 for SF3A, and *Schizosaccharomyces pombe* Mae1 for SF3B. **b**, Structure-based sequence alignment of TehA from *H. influenzae* (HiTehA) and SLAC1 from *A. thaliana* (AtSLAC1). The TehA structure has been used to restrict sequence gaps to inter-helical segments. Coils above residues define the extent of the HiTehA helical segments; red letters mark residue identities; red boxes are drawn for residues that are >95% identical within the plant subfamily SF1A for AtSLAC1 or within the TehA subfamily for HiTehA; red diamonds mark HiTehA residues that line the central pore; and the coloured bar below residues encodes ConSurf sequence variability[47] for the SF1 family of 204 non-redundant proteins.

as exfoliative toxins; and family SF3 comprises the fungal Mae1 and Ssu1 proteins and their archaeal or bacterial homologues, respectively. SLAC family SF1 has three large subfamilies: the plant SLAC and SLAH proteins are in subfamily SF1A, closest bacterial homologues are in SF1B, and the TehA homologues are in SF1C (Fig. 1a). The other families also

divide into subfamilies as detailed in Supplementary Table 1, and family SF1 is divided into sub-subfamilies (Supplementary Table 2). Two pertinent SF1 sequences are aligned in Fig. 1b.

We used a structural genomics approach to obtain structural information, testing expression and purification for forty-three bacterial and
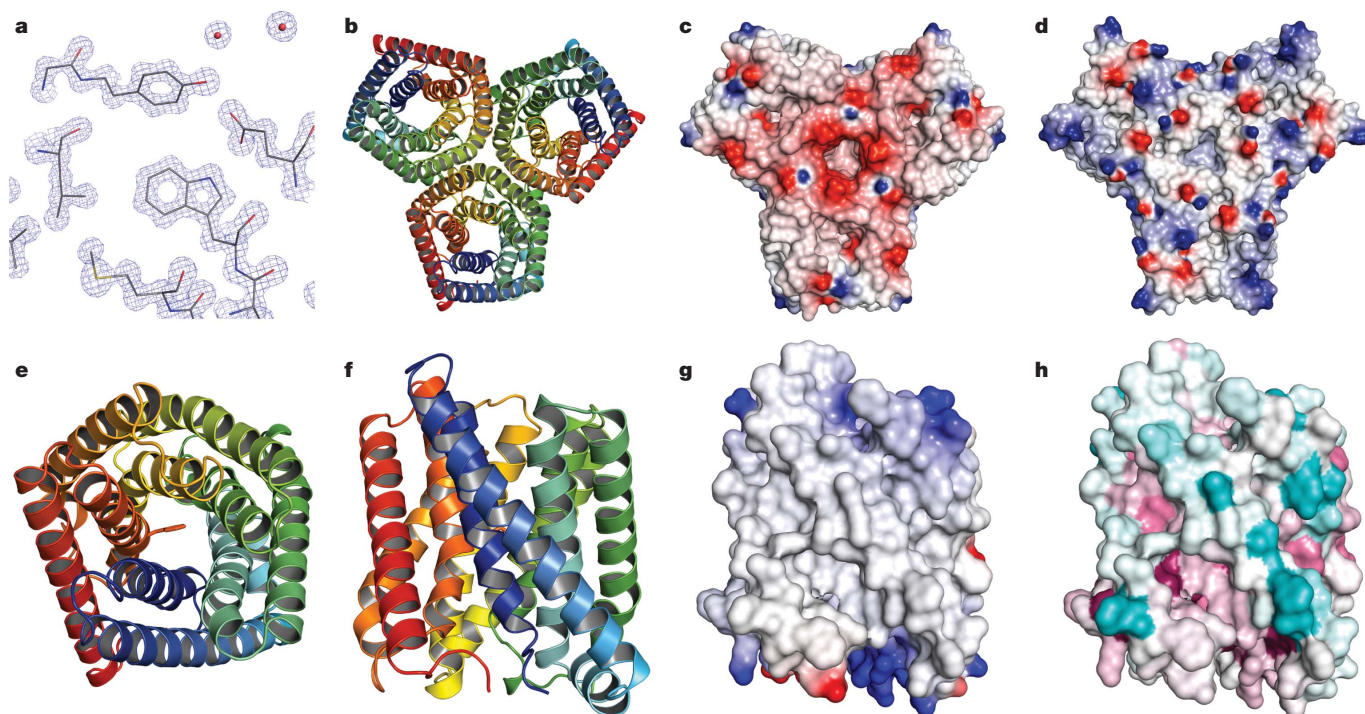
Figure 2 | Crystal structure of HiTehA and homology model of AtSLAC1. a, Electron density distribution from the HiTehA crystal structure at 1.2 Å resolution. The map has $(2F_o - F_c)$ coefficients based on the superimposed model. Contours are at 2.5σ. **b**, Ribbon diagram of the HiTehA trimer. Each protomer is coloured spectrally from blue at its N terminus to red at its C terminus. **c**, DelPhi[48] electrostatic potential at the extracellular surface. Electronegative and electropositive potential are coloured in degrees of red and blue saturation, respectively. **d**, Electrostatic potential[48] at the intracellular surface. **e**, Ribbon diagram of an HiTehA protomer viewed from outside the membrane. The ribbon is coloured spectrally as in **b**. **f**, Ribbon diagram of an HiTehA protomer viewed from within the membrane, 90° from the view of **e** and with the cytoplasm below. **g**, Surface of a homology model of AtSLAC1, viewed as in **f**, and coloured by electrostatic potential[48]. **h**, Surface of AtSLAC1 as in **g**, but coloured by sequence variability[47] as in Fig. 1b.

archaeal likely homologues, assaying for detergent choice and stability on eight of these, finding two with appropriate profiles by size-exclusion chromatography, and obtaining suitable crystals for one. This protein, TehA from *H. influenzae* (HiTehA), was found to be trimeric both by size-exclusion multi-angle light-scattering (SEC–MALS) measurements and by chemical cross-linking. When solubilized in β-octylglucoside, HiTehA crystallized in space group $R3$ with $a = b = 96.01$ Å and $c = 136.27$ Å. Each asymmetric unit contains one subunit and 65% solvent. The structure was solved by selenomethionyl (SeMet) SAD phasing, ultimately at 1.50 Å resolution (Supplementary Table 3 and Supplementary Fig. 1), and then refined at 1.20 Å resolution (Fig. 2a) to R/R_{free} values of 14.1%/16.0% for a model that includes ordered residues 6–313, 213 water molecules and four detergent molecules (Supplementary Table 4).

The crystal structure has TehA trimers aligned with three-fold axes of the lattice (Fig. 2b). Subunits are tightly associated, burying 8,947 Å² of total surface area within trimer interfaces. The electrostatic potential surface is largely negative on the extracellular surface (Fig. 2c) and largely positive on the cytoplasmic surface (Fig. 2d). The membrane orientation is specified experimentally from GFP tagging of *E. coli* TehA[24]. Each TehA protomer has ten transmembrane helices, as predicted; however, the fold is novel. Tandemly repeated helical hairpins are arranged with quasi-five-fold symmetry (Fig. 2e and Supplementary Fig. 2). Extracellular inter-helix loops are short (2–5 residues), whereas intracellular inter-helix connections are longer, including a nine-residue helix $H_{2,3}$ between transmembrane helix 2 (TM_2) and TM_3 (Fig. 2f). An inner pentad of outwardly directed, TM_{odd}, helices creates an apparent pore through each protomer perpendicular to the putative membrane plane. TM_{even} helices from the five hairpins surround the inner pore and make an outer layer.

Homology model for plant SLAC1

Arabidopsis SLAC1 (AtSLAC1) is substantially similar to bacterial homologues, notably HiTehA (Fig. 1b). All HiTehA transmembrane helices are fully aligned to predicted SLAC1 transmembrane helices, but there are short inter-helical gaps (1–5 residues) in all five extracellular loops and in two of the intracellular loops. The transmembrane domain of AtSLAC1 (residues 188–504) aligns to HiTehA with 19% sequence identity and with a PSI-BLAST E-value of 3×10^{-22}. For comparison, and in keeping with the family tree (Fig. 1a), the transmembrane domain of AtSLAC1 shares sequence identities of 76% with rice SLAC1, 41% with *Arabidopsis* SLAH1, 25% with an SL1B homologue from *Halorhodospira halophila*, 11% with *S. cerevisiae* Ssu1 and 9% with *S. pombe* Mae1. A conceptual model with the AtSLAC1 sequence transposed onto the HiTehA helices sufficed to guide most of our mutational tests, but a detailed AtSLAC1 homology model helped to refine our ideas. Electrostatic potential and surface variability are plotted onto the surface of this model (Fig. 2g, h).

The most remarkable feature of the TehA structure and corresponding SLAC1 model is the central pore through each protomer. As is the case for acetylcholine receptors[25], the SLAC1 pore is formed by five helices, but the SLAC1 helices come from one protein molecule rather than five. The SLAC1 pore has a relatively uniform diameter of approximately 5 Å across nearly five helical turns (Supplementary Fig. 3), except for a pronounced constriction in the middle of the membrane (Fig. 3a) where the pore is occluded by the side chain of Phe 450 (Phe 262 in HiTehA). This residue is the only absolutely conserved amino acid residue of the SLAC1 family. The pore is lined with highly conserved (86% identity among five SLAC1 orthologues; 32% identity between AtSLAC1 and HiTehA) and generally hydrophobic residues (Figs 1b and 3b, c and Supplementary Fig. 4). Despite this
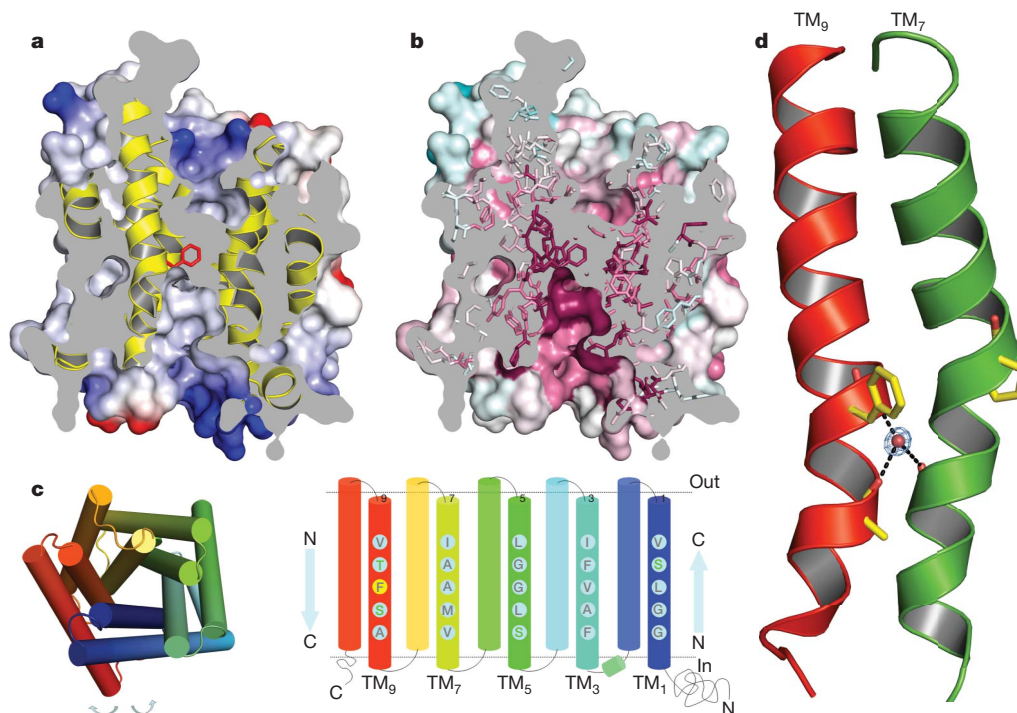
Figure 3 | Putative structure of the SLAC1 conductance pore. **a**, Cross-section through the homology model of AtSLAC1. The model is viewed as in Fig. 2g, with the electrostatic potential[48] shown on the external surface of the molecular envelope. The side chain of Phe 450 is shown as a stick model (red) on the backbone ribbon, coloured yellow. **b**, Cross-section as in **a**, but coloured by surface conservation[47] as in Fig. 2h. **c**, Pore-lining residues in the SLAC1 homology model. A cylinder model (left), spectrally coloured as in Fig. 2b, provides a key for viewing the rolled-open structure (right) with pore-lining residues of AtSLAC1 shown on the TM_{odd} helices. The yellow background highlights conserved Phe 450 and green letters show hydroxyl-bearing residues. **d**, Ribbon diagrams of HiTehA TM_9 (left) and TM_7 (right) viewed from within the conductance pore. The side chains of Pro 207 and Phe 262 are shown as well as a kink-stabilizing water HOH25 that is coordinated by the NH of Gly 263 and by C=O groups of Gly 202 and Ala 259. Density contours are shown for the water molecule.

hydrophobicity, the electrostatic potential on the pore surface is polarized (Fig. 3a), presumably due to an invaginated shape adjacent to charged residues outside the membrane. The generally electropositive character of the cytoplasmic surface probably contributes to anion efflux.

Kinks in the pore helices contribute to formation of a relatively constant pore diameter across the membrane. Four of the five HiTehA inner helices have centrally located proline residues, which necessarily generate kinks, and TM_9 is kinked at a backbone-coordinated water molecule (Fig. 3d). Proline replaces Gly 263 in all SF1A and SF1B relatives, including Pro 451 of AtSLAC1. This water-displacing change is isostructural (Y.-H.C., L.H., S.A.S. and W.A.H., unpublished data). Centrally located proline residues also prevail in TM_3, TM_5 and TM_7 across the SF1 proteins. The outer helices are longer and straighter, but more inclined. The only outer-helix proline kink of HiTehA is in TM_6 at the trimer three-fold axis.

Two of the gene-identifying mutations in *Arabidopsis* SLAC1 are selected point mutations; others are disruptive transfer DNA (T-DNA) insertions[4]. In the AtSLAC1 homology model, the *slac1-2* (G194D) mutation[3] points into the pore from TM_1 and can be accommodated structurally, whereas the *slac1-1* (S456F) mutation[4] points away from the pore six residues after pore-blocking Phe 450 on TM_9 and would be expected to be disruptive. Residue Ser 456 interacts with outer-helix TM_{10} in the homology model, and the phenyl bulk from S456F would not fit. Position 456 has alanine in HiTehA and also in 66% of all 204 SF1 homologues, whereas another 27% have threonine or serine (as in all SLAC1 channels); phenylalanine never occurs at position 456 among all 814 SLAC superfamily members. Position 194 has glycine in 58% of SF1 homologues and alanine in another 22%. Residue

Asp 194, which never occurs naturally, would block the pore and is expected to repel anions.

Mutational tests of channel function

Mutational studies corroborate the hypothesis that a TehA-based SLAC1 model is appropriate. First, as discussed above, the *slac1-1* mutant (S456F) is expected to be structurally disruptive—and indeed it is inactive in guard cells[4]—and the *slac1-2* mutation (G194D) is expected to block the pore, and we show below that this variant is also inactive. We have also shown that the introduction of SLAC1-conserved proline residues into HiTehA (A208P/G263P) is accommodated isomorphously (Y.-H.C., L.H., S.A.S. and W.A.H., unpublished data). Moreover, as shown below, channel conductance properties of several mutants are similar for AtSLAC1 and HiTehA.

To examine characteristics of the SLAC1 channel in light of the structural model, we performed electrophysiological tests of membrane currents from voltage-clamped *Xenopus* oocytes after injection of wild-type or mutant AtSLAC1 or HiTehA cRNAs. We observed modest-sized Cl^- currents with wild-type AtSLAC1 cRNA, as found previously[6,7], but did not detect any Cl^- current after injection of wild-type HiTehA cRNA. We found that SLAC1 Cl^- conductance was enhanced when the OST1 kinase cRNA was co-injected with SLAC1 cRNA, but only to the levels found by ref. 6 and not to the much higher levels found by ref. 7 with OST1 physically connected to SLAC1 by split YFP linkage. Consistent with the structural evidence that Phe 262 blocks the HiTehA pore, removal of the phenyl group in HiTehA F262A or in the homologous AtSLAC1 F450A mutant resulted in very large Cl^- currents relative to wild-type levels, and the SLAC1 currents were now less enhanced by the presence of OST1
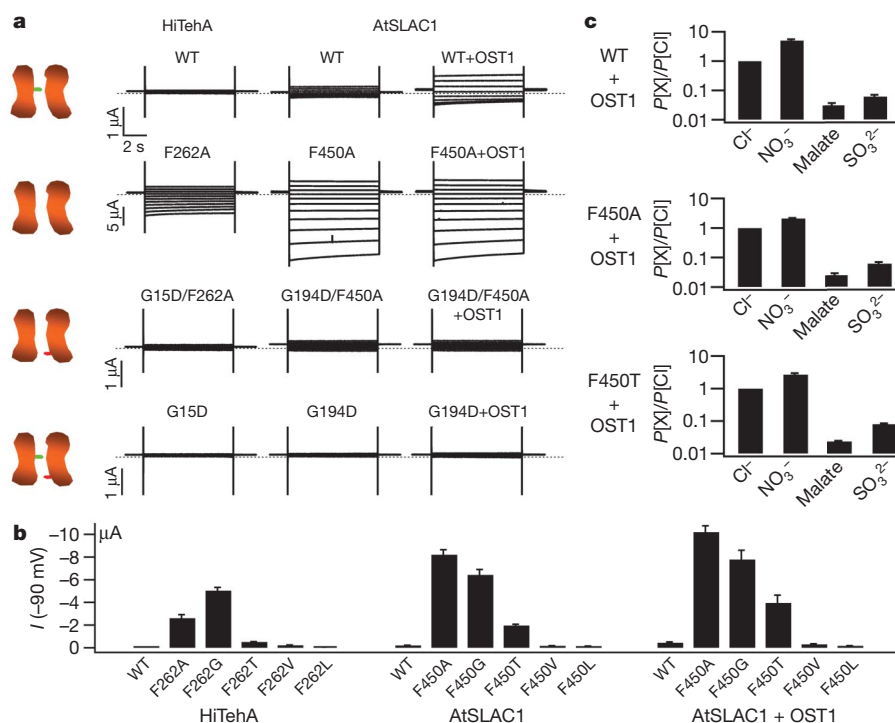
Figure 4 | Ionic conductance measurements. a, Representative microelectrode voltage-clamp current traces from oocytes injected with various channel cRNAs. Left column, oocytes injected with cRNAs encoding wild-type HiTehA channels (WT), or F262A, G15D/F262A or G15D mutants. Middle and right columns, oocytes injected with cRNAs for wild-type AtSLAC1, or F450A, G194D/F450A or G194D mutants, with or without co-injection of AtOST1. Dotted lines represent zero current levels. The extracellular solution contained 30 mM CsCl. Schematic icons at the far left show the phenyl gate (green) of wild-type channels and/or the aspartyl barrier (red) of the G194D or G15D mutants. **b,** Effects of gating residue mutations. Mean chloride currents, measured at −90 mV, are shown comparing wild-type HiTehA with its mutant series F262A, F262G, F262T, F262V, F262L and wild-type AtSLAC1 with its corresponding series F450A, F450G, F450T, F450V, F450L, both alone and co-expressed with AtOST1. Full *I–V* relations are shown in Supplementary Fig. 5. **c,** Effect of substitutions for gating residue Phe 450 on relative AtSLAC1 anion permeabilities. Relative permeabilities ($P[X]/P[Cl]$) for chloride, nitrate, sulphite and malate of wild-type, F450A and F450T SLAC1 channels were measured from the change in current reversal potential with Cl^- or anion X^- as the sole permeant anion in the bath solution (see Methods and Supplementary Table 6).

(Fig. 4a). The tempting interpretation of a constitutively opened gate in F450A will require validation with appropriately analysed single-channel recordings[26]. In keeping with the *slac1-2* phenotype[3], neither the functionally impaired AtSLAC1 G194D nor its HiTehA G15D homologue showed any substantial conductance; moreover, consistent with pore blockage by Asp 194 in *slac1-2*, the large conductances of HiTehA F262A and AtSLAC1 F450A were abolished in the double mutants AtSLAC1 G194D/F450A and HiTehA G15D/F262A (Fig. 4a). Here again, the effects in SLAC1 were independent of OST1.

We also tested the conductance characteristics for a series of AtSLAC1 F450X substitution mutants—F450A, F450G, F450T, F450V and F450L—and for the corresponding HiTehA F262X series—F262A, F262G, F262T, F262V and F262L (Fig. 4b, Supplementary Fig. 5 and Supplementary Table 5). Findings from the two series are roughly parallel; in particular, the alanine and glycine substitutions lead to large currents for both and in comparison to the others. There are distinctions, of course, including generally higher conductances for AtSLAC1 over HiTehA and less conductance of F262T TehA compared to F450T SLAC1. It is also noteworthy that OST1 activation is very muted for the F450A, F450G and F450L mutants, which is consistent with SLAC1 gating at Phe 450.

Crystal structures were also determined for several of the HiTehA mutant variants (Supplementary Table 4). The structures of F262A (1.15 Å), F262V (1.60 Å), F262L (1.65 Å) and G15D (1.50 Å) are all essentially isomorphous with the wild-type TehA structure, with changes localized primarily at the sites of mutation; the same is true for the double mutations of F262A/G15D, F262G/G15D and A208P/G263P (Y.-H.C., L.H., S.A.S. and W.A.H., unpublished data). The F262A structure has a wide-open pore (Fig. 5a) with a relatively

uniform pore diameter of ∼5 Å through ∼30 Å across the membrane (Supplementary Fig. 3), whereas G15D has a doubly occluded pore (Fig. 5b). The pores of other mutant variants are consistent with the sizes of constrictive residues and with the observed conductances.

Gating and activation

The crystal structures of TehA and its mutant variants when taken together with the functional studies in *Xenopus* oocytes point to a crucial role for Phe 450 in gating of the SLAC1 anion channel. Conservation of this residue across the SLAC1 family implies functional importance. The occlusion of the pore by the presence of Phe 262 in the structure of wild-type TehA and the openness of the pore upon substitution of phenylalanine by alanine in the structure of the F262A mutant provides physical evidence for a gating role of this residue. This interpretation is supported by the correlated conductance characteristics from variants of the AtSLAC1 and HiTehA channels (Fig. 4b and Supplementary Fig. 5). Although these observations may suffice for placing the gate within the channel pore, they do not by themselves suggest a mechanism for gating in response to physiological stimuli. Some insight does come from conformational details defined at high resolution.

One important structural clue is that the side chain of Phe 262 is in a high-energy conformation in the HiTehA structure, with χ_1/χ_2 at $-160°/-4°$. Although χ_1 is in a preferred *trans* conformation, the phenyl ring is restricted by contacts with Val 210 and Leu 18 to a χ_2 value near to $0°$ rather than to the preferred $90°$ orientation (Fig. 5c). Further evidence that Phe 262 is restrained from local equilibrium comes from shifts observed in crystal structures of the F262A, F262V, F262L and F262G/G15D variants, which all show consistent
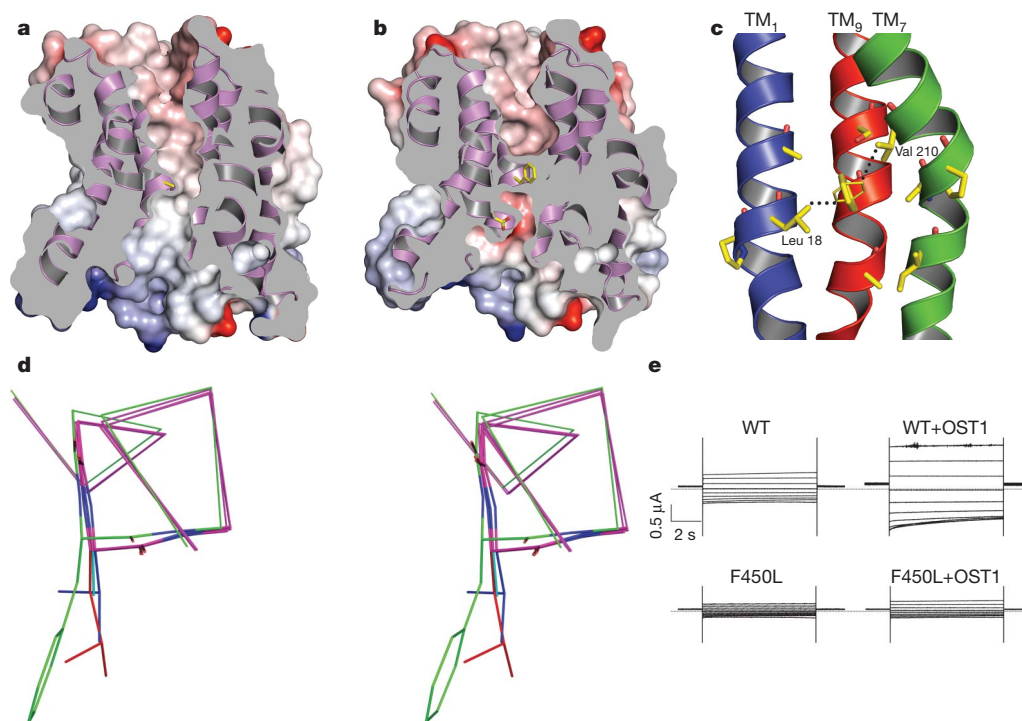
Figure 5 | Structural features at the SLAC1 homologue phenylalanine gate. **a**, **b**, Cross-sections through the conductance pores of HiTehA F262A and HiTehA G15D. The view and presentations are as in Fig. 3a, except that helices are coloured purple. **c**, Molecular basis for conformational strain in gating residue Phe 262 of HiTehA. Helices TM$_1$ (left), TM$_9$ (centre) and TM$_7$ (right) are viewed from within the pore and presented as ribbon diagrams with selected side chains drawn in stick representation. The local low-energy conformation for the phenyl ring ($\chi_2 = 90°$) is shown in thin lines with short contacts indicated by dashed lines: Leu 18 C$_{\delta2}$ to Phe 262 C$_{\delta1}$, distance = 2.4 Å; Val 210 C$_{\gamma2}$ to Phe 262 C$_{\varepsilon2}$, distance = 2.8 Å. **d**, Conformational shifts consequent to

release of strain in gating residue Phe 262. Cα backbone structures of F262A, F262V and F262L HiTehA are superimposed onto wild-type HiTehA. Residues 258–266 from this superposition are drawn in stereo, with all backbone atoms shown for peptides 262 ± 1 but only Cα atoms shown otherwise. The wild-type backbone and phenyl group are green; all other backbones are magenta; side chains of Ala 262, Val 262 and Leu 262 are cyan, blue and red, respectively; oxygen-directed bonds are red and nitrogen-directed bonds are blue. **e**, Representative microelectrode voltage-clamp current traces from oocytes injected with wild-type (WT) or F450L AtSLAC1 cRNA. Experimental conditions and displays are as in Fig. 4a.

backbone movements that displace Cβ(262) by 0.47–0.78 Å (Fig. 5d). By contrast, Leu 262 in F262L is in a preferred trans/gauche$^+$ conformation at $\chi_1/\chi_2 = 177°/63°$ as is Val 262 in F262V at $\chi_1 = -176°$. What might control activation of HiTehA is unclear, but for AtSLAC1 activation is by OST1 phosphorylation[6,7]. The molecular consequences of OST1 phosphorylation of SLAC1 remain unknown, but it is plausible that associated shifts in pore-helix orientations would unlatch Phe 450 in SLAC1 from a TehA-like restrained orientation. By analogy with Leu 262 in F262L, we expect a preferred rotameric state for Leu 450 in the AtSLAC1 F450L variant. Thus, the lack of appreciable OST1 activation of conductance in the AtSLAC1 F450L variant (Fig. 5e) might be explained by the lack of a restraining latch, whereby the channel remains closed despite OST1 activation. Puzzles certainly remain, because OST1 does substantially activate AtSLAC1 F450T and F450V, which like HiTehA F262V should also be unrestrained; presumably, activating adjustments widen the pore enough for ion permeation past threonine and valine but not leucine.

Phosphorylation sites have been discovered in the N- and C-terminal tails of AtSLAC1[6,7,27] (179 and 51 residues long, respectively), but these alone cannot explain OST1 activation of SLAC1. First, SLAH1, which fully complements the *slac1-1* mutation, does not have these cytoplasmic tails. Second, although OST1 phosphorylation of Ser 120 in the N-terminal tail is necessary for SLAC1 activation, it is not sufficient[7]. Thus, we surmise that direct phosphorylation of the SLAC1 transmembrane domain must be critical, and SLAC1 has four conserved Ser/Thr candidates in its cytoplasmic loops. Moreover, SLAC1 proteins have proline-mediated kinks at the putative Phe 450 gate in helix TM$_9$, and also in adjacent helix TM$_7$; these features may have a role in phosphorylation-driven unlatching of the Phe 450 gate in SLAC1.

Ion selectivity and discrimination

Our studies of SLAC1 channel relative ion permeabilities, based on measurements of current reversal potential, are consistent with earlier work demonstrating that AtSLAC1 conducts anions but not cations and is selective among anions, with greater permeability for nitrate than for chloride (as in *Vicia faba* guard cell protoplasts[26]) and much reduced permeability for malate, bicarbonate or sulphate[6,7]. We also find that SLAC1 has little permeability for sulphite. Additionally, we find that wild-type SLAC1, F450A and F450T all have similar relative permeabilities to chloride, sulphite and malate, despite having widely different conductance levels, but the gating mutants do show small but significant decreases in their nitrate to chloride permeability ratios (Fig. 4c and Supplementary Table 6).

The relative insensitivity of anion permeability to gating residue changes indicates that selectivity for these anions may occur away from the central constriction at the channel gate. To some extent, ionic discrimination must depend on pore geometry; thus, an organic anion such as malate may be simply too large to pass through the 5-Å-wide pore. Although the SLAC1 pore is lined largely with hydrophobic side chains (Fig. 3c), it also has a few hydroxyl groups from serine and threonine residues (16%) whose electropositive hydrogen atoms may facilitate conductance. Most notably, the electrostatic potential within the AtSLAC1 pore is electropositive throughout (Fig. 3a). This polarization, promoted by charges on extra-membranous loops, no doubt contributes significantly towards discrimination against cations.

The relative anion permeability sequence of SLAC1 determined by us and others, $I^- > NO_3^- > Br^- > Cl^-$ (refs 6, 7), corresponds to selectivity sequence 1 compiled by ref. 28 for a range of anion-selective proteins. This sequence correlates inversely with the hydration energies of monovalent anions—anions with a lower hydration energy have a greater channel permeability. It is thought to be generated in proteins with weak, low-field-strength anion-binding sites, where selectivity is largely determined by the energetic cost of anion dehydration. These selectivity

results are thus consistent with the SLAC1 structure, where the pore lacks any obvious anion-binding site.

Distinctiveness of the SLAC1 channel

SLAC1 anion channels are entirely novel in structure and, apparently, in the mechanism for ion conductance. The best characterized of anion channels belong to the CLC family of Cl^- channels and transporters[29–31]. CLC channels have an altogether different architecture from the SLAC1 channel, and the mechanism for selectivity is also very different. Bacterial CLC transporters bind halide ions at three sites in a highly constricted pore[29]. By contrast, the SLAC1 pore has a relatively uniform diameter across the membrane, except where closed by the gating phenyl group, and we do not find discrete ion binding sites. CLC selectivity is governed by specific residues surrounding these binding sites[29,31]. The anion selectivity sequence for the CLC channels of $Cl^- > Br^- > NO_3^- > I^-$, opposite of that in SLAC1, is consistent with the high-field-strength anion-binding sites in CLC channels[28]. Interestingly, as for AtSLAC1, an *Arabidopsis* CLCa channel also preferentially transports nitrate ions[32], and an *E. coli* CLC channel is converted to a preference of nitrate when a generally conserved serine at the central site is substituted with proline, as in AtCLCa[31].

SLAC1 also differs radically from other structurally characterized anion channels and transporters. These include the VDAC1 voltage-gated anion channel from mitochondrial outer membranes, which has a porin-like β-barrel structure[33–35], and a light-driven halorhodopsin chloride pump, which has a transmembrane conductance pathway similar to that of the proton-pumping pore of bacteriorhodopsin[36]. Although its channel structure is still only known by homology to other ABC transporters, CFTR is another obviously distinct chloride channel[37]. Cys-loop receptors also include anion channels[38], and these are similar to SLAC1 in having five-helix pores[25], but here selectivity is governed by charged groups at the entrance to the pore, which distinguish the anion-selective GABA$_A$ and glycine receptors from the cation-selective acetylcholine and serotonin 5HT3 receptors[38]. Finally, recently identified *TMEM16A* genes for calcium-activated chloride channels[39–41] seem to encode an 8-transmembrane protein that is again distinct from SLAC1.

Stomatal guard cells show both rapidly activated (R-type) and slow (S-type) anion channel activity[42]. Although *slac1* guard cells have very defective S-type activity, their R-type currents are normal[4]. Guard cell protoplasts from the *slac1-2* mutant abnormally accumulate Cl^-, K^+, malate and fumarate[3], whereas SLAC1 shows negligible malate conductance[7]. As for SLAC1-associated K^+ movements, other channels or transporters must be responsible for SLAC1-associated malate movements. Recent studies indicate that AtALMT12, an aluminium-activated malate transporter (ALMT) family member, is a malate-dependent R-type anion channel[43] needed for stomatal closure[44].

Conclusions

We find that many functional properties of the plant SLAC1 anion channel are explained well by the structure of a previously uncharacterized bacterial TehA protein that has been associated with tellurite resistance. SLAC1 and TehA belong to distinct subfamilies within one branch of a larger SLAC1 superfamily, but AtSLAC1 and HiTehA are sufficiently similar (19% sequence identity) that the SLAC1 homology model is predictive for function, including a verified placement of the identifying *slac1-2* mutation G194D and a phenylalanine gate. Two questions that remain concern the structural change that activating phosphorylation elicits in SLAC1, and the biochemical role of the TehA homologues in bacteria. Elsewhere (Y.-H.C., L.H., S.A.S. and W.A.H., unpublished data), we examine the functional and structural properties of TehA in bacteria, showing that it is an anion channel, although actually not conferring tellurite resistance, and identifying a mutant variant with properties indicative of an activated state. Thus,

SLAC1 and TehA probably represent a large family of selective anion channels controlled by environmental stimuli.

METHODS SUMMARY

NYCOMPS pipeline procedures[45] were used to identify prokaryotic homologues of *E. coli* TehA and to test for suitability for detergent solubilization and purification. *H. influenzae* TehA was stably purified and crystals that diffracted beyond 1.1 Å spacings were obtained at 4 °C from protein in β-octylglucoside detergent at a range of pH values buffered from 5.2 to 10.2 and with PEG600 or PEG400 as the precipitant. The structure of HiTehA was solved from SeMet SAD measurements, initially at 2.0 Å but later extended to 1.5 Å resolution, and then refined at 1.20 Å resolution for wild-type HiTehA. A thorough bioinformatic analysis showed that TehA and plant SLAC1 proteins are close relatives in a family distinct from other homologues. Mutant variants were prepared by site-directed mutagenesis in both HiTehA and AtSLAC1. Corresponding cRNAs were injected into *Xenopus* oocytes, and conductance properties were studied in whole-cell voltage-clamp recordings.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 28 March; accepted 10 September 2010.

1. Hetherington, A. M. & Woodward, F. I. The role of stomata in sensing and driving environmental change. *Nature* **424,** 901–908 (2003).
2. Sirichandra, C., Wasilewska, A., Vlad, F., Valon, C. & Leung, J. The guard cell as a single-cell model towards understanding drought tolerance and abscisic acid action. *J. Exp. Bot.* **60,** 1439–1463 (2009).
3. Negi, J. et al. CO$_2$ regulator SLAC1 and its homologues are essential for anion homeostatsis in plant cells. *Nature* **452,** 483–486 (2008).
4. Vahisalu, T. et al. SLAC1 is required for plant guard cell S-type anion channel function in stomatal signalling. *Nature* **452,** 487–491 (2008).
5. Saji, S. et al. Disruption of a gene encoding C4-dicarboxylate transporter-like protein increases ozone sensitivity through deregulation of the stomatal response in *Arabidopsis* thaliana. *Plant Cell Physiol.* **49,** 2–10 (2008).
6. Lee, S. C., Lan, W., Buchanan, B. B. & Luan, S. A protein kinase-phosphatase pair interacts with an ion channel to regulate ABA signaling in plant guard cells. *Proc. Natl Acad. Sci. USA* **106,** 21419–21424 (2009).
7. Geiger, D. et al. Activity of guard cell anion channel SLAC1 is controlled by drought-stress signaling kinase-phosphatase pair. *Proc. Natl Acad. Sci. USA* **106,** 21425–21430 (2009).
8. Schroeder, J. I. & Hagiwara, S. Cytosolic calcium regulates ion channels in the plasma membrane of *Vicia faba* guard cells. *Nature* **338,** 427–430 (1989).
9. Mustilli, A., Merlot, S., Vavasseur, A., Fenzi, F. & Giraudat, J. *Arabidopsis* OST1 protein kinase mediates the regulation of stomatal aperture by abscisic acid and acts upstream of reactive oxygen species production. *Plant Cell* **14,** 3089–3099 (2002).
10. Leung, J. et al. *Arabidopsis* ABA response gene *ABI1*: features of a calcium-modulated protein phosphatase. *Science* **264,** 1448–1452 (1994).
11. Meyer, K., Leube, M. P. & Grill, E. A protein phosphatase 2C involved in ABA signal transduction in *Arabidopsis thaliana*. *Science* **264,** 1452–1455 (1994).
12. Ma, Y. et al. Regulators of PP2C phosphatase activity function as abscisic acid sensors. *Science* **324,** 1064–1068 (2009).
13. Park, S. et al. Abscisic acid inhibits type 2C protein phosphatases via the PYR/PYL family of START proteins. *Science* **324,** 1068–1071 (2009).
14. Melcher, K. et al. A gate-latch-lock mechanism for hormone signalling by abscisic acid receptors. *Nature* **462,** 602–608 (2009).
15. Miyazono, K. et al. Structural basis of abscisic acid signalling. *Nature* **462,** 609–614 (2009).
16. Fujii, H. et al. In vitro reconstitution of an abscisic acid signalling pathway. *Nature* **462,** 660–664 (2009).
17. Yin, P. et al. Structural insights into the mechanism of abscisic acid signaling by PYL proteins. *Nature Struct. Mol. Biol.* **16,** 1230–1236 (2009).
18. Nishimura, N. et al. PYR/PYL/RCAR family members are major in-vivo ABI1 protein phosphatase 2C-interacting proteins in *Arabidopsis*. *Plant J.* **61,** 290–299 (2010).
19. Grobler, J., Bauer, F., Subden, R. E. & van Vuuren, H. J. The *MAE1* gene of *Schizosaccharomyces pombe* encodes a permease for malate and other C4 dicarboxylic acids. *Yeast* **11,** 1485–1491 (1995).
20. Park, H. & Bakalinsky, A. T. *SSU1* mediates sulphite efflux in *Saccharomyces cerevisiae*. *Yeast* **16,** 881–888 (2000).
21. Léchenne, B. et al. Sulphite efflux pumps in *Aspergillus fumigatus* and dermatophytes. *Microbiology* **153,** 905–913 (2007).
22. Walter, E. G., Weiner, J. H. & Taylor, D. E. Nucleotide sequence and overexpression of the tellurite-resistance determinant from the IncHII plasmid pHH1508a. *Gene* **101,** 1–7 (1991).
23. Taylor, D. E., Hou, Y., Turner, R. J. & Weiner, J. H. Location of a potassium tellurite resistance operon (*tehA tehB*) within the terminus of *Escherichia coli* K-12. *J. Bacteriol.* **176,** 2740–2742 (1994).
24. Daley, D. O. et al. Global topology analysis of the *Escherichia coli* inner membrane proteome. *Science* **308,** 1321–1323 (2005).
25. Unwin, N. Refined structure of the nicotinic acetylcholine receptor at 4 Å resolution. *J. Mol. Biol.* **346,** 967–989 (2005).
26. Schmidt, C. & Schroeder, J. I. Anion selectivity of slow anion channels in the plasma membrane of guard cells (large nitrate permeability). *Plant Physiol.* **106,** 383–391 (1994).
27. Vahisalu, T. et al. Ozone-triggered rapid stomatal response involves the production of reactive oxygen species, and is controlled by SLAC1 and OST1. *Plant J.* **62,** 442–453 (2010).
28. Wright, E. M. & Diamond, J. M. Anion selectivity in biological systems. *Physiol. Rev.* **57,** 109–156 (1977).
29. Dutzler, R., Campbell, E. B. & MacKinnon, R. Gating the selectivity filter in ClC chloride channels. *Science* **300,** 108–112 (2003).
30. Accardi, A. & Miller, C. Secondary active transport mediated by a prokaryotic homologue of ClC Cl$^-$ channels. *Nature* **427,** 803–807 (2004).
31. Picollo, A., Malvezzi, M., Houtman, J. C. & Accardi, A. Basis of substrate binding and conservation of selectivity in the CLC family of channels and transporters. *Nature Struct. Mol. Biol.* **16,** 1294–1301 (2009).
32. De Angeli, A. et al. The nitrate/proton antiporter AtCLCa mediates nitrate accumulation in plant vacuoles. *Nature* **442,** 939–942 (2006).
33. Hiller, S. et al. Solution structure of the integral human membrane protein VDAC-1 in detergent micelles. *Science* **321,** 1206–1210 (2008).
34. Bayrhuber, M. et al. Structure of the human voltage-dependent anion channel. *Proc. Natl Acad. Sci. USA* **105,** 15370–15375 (2008).
35. Ujwal, R. et al. The crystal structure of mouse VDAC1 at 2.3 Å resolution reveals mechanistic insights into metabolite gating. *Proc. Natl Acad. Sci. USA* **105,** 17742–17747 (2008).
36. Kouyama, T. et al. Crystal structure of the light-driven chloride pump halorhodopsin from *Natronomonas pharaonis*. *J. Mol. Biol.* **396,** 564–579 (2010).
37. Gadsby, D. C., Vergani, P. & Csanády, L. The ABC protein turned chloride channel whose failure causes cystic fibrosis. *Nature* **440,** 477–483 (2006).
38. Miller, P. S. & Smart, T. G. Binding, activation and modulation of Cys-loop receptors. *Trends Pharmacol. Sci.* **31,** 161–174 (2010).
39. Yang, Y. D. et al. TMEM16A confers receptor-activated calcium-dependent chloride conductance. *Nature* **455,** 1210–1215 (2008).
40. Caputo, A. et al. TMEM16A, a membrane protein associated with calcium-dependent chloride channel activity. *Science* **322,** 590–594 (2008).
41. Schroeder, B. C., Cheng, T., Jan, Y. N. & Jan, L. Y. Expression cloning of TMEM16A as a calcium-activated chloride channel subunit. *Cell* **134,** 1019–1029 (2008).
42. Schroeder, J. I. & Keller, B. U. Two types of anion channel currents in guard cells with distinct voltage regulation. *Proc. Natl Acad. Sci. USA* **89,** 5025–5029 (1992).
43. Meyer, S. et al. AtALMT12 represents an R-type anion channel required for stomatal movement in *Arabidopsis* guard cells. *Plant J.* **63,** 1054–1062 (2010).
44. Sasaki, T. et al. Closing plant stomata requires a homolog of an aluminum-activated malate transporter. *Plant Cell Physiol.* **51,** 354–365 (2010).
45. Punta, M. et al. Structural genomics target selection for the New York consortium on membrane protein structure. *J. Struct. Funct. Genomics* **10,** 255–268 (2009).
46. Papadopoulos, J. S. & Agarwala, R. COBALT: constraint-based alignment tool for multiple protein sequences. *Bioinformatics* **23,** 1073–1079 (2007).
47. Landau, M. et al. ConSurf 2005: the projection of evolutionary conservation scores of residues on protein structures. *Nucleic Acids Res.* **33,** W299–W302 (2005).
48. Rocchia, W. et al. Rapid grid-based construction of the molecular surface for both molecules and geometric objects: applications to the finite difference Poisson-Boltzmann method. *J. Comput. Chem.* **23,** 128–137 (2002).

Author Contributions Y.-H.C., L.H., M.P., R.B., B.H., B.K. and J.L. performed experiments; Y.-H.C., L.H., B.R., S.A.S. and W.A.H. analysed data; Y.-H.C., L.H., S.A.S. and W.A.H. prepared the manuscript.

METHODS

Selection of target sequences. TehA from *E. coli* (EcTehA) was a centrally selected sequence for expansion as a NYCOMPS targeted family. Target selection criteria used at NYCOMPS are described in detail elsewhere[45]. Briefly, the EcTehA sequence was run against a data set of about 40,000 predicted α-helical integral membrane protein sequences from prokaryotic genomes (NYCOMPS98 data set[45]) using PSI-BLAST[49]. Sequences that matched EcTehA with an *E*-value lower than 10^{-3} in an alignment extending over at least 50% of both predicted transmembrane regions and passing our post-seed-expansion filtering criteria[45] were passed to the protein production pipeline.

Protein expression screening. Full-length homologues from the following 38 species, including 2 sequences each from 5 of these, were amplified from genomic DNA by PCR: *Thermoplasma acidophilum, Lactococcus lactis* subsp. lactis Il1403, *Streptococcus pyogenes* M1 GAS, *Streptococcus pneumoniae* TIGR4, *Haemophilus influenzae* Rd KW20, *Methanocaldococcus jannaschii* DSM 2661 (2), *Escherichia coli* K12, *Salmonella typhimurium* LT2, *Clostridium perfringens* ATCC 13124, *Xanthomonas campestris* pv. campestris strain ATCC 33913, *Streptococcus agalactiae* 2603V/R, *Shewanella oneidensis* MR-1, *Streptococcus mutans* UA159, *Archaeoglobus fulgidus* DSM 4304, *Vibrio parahaemolyticus* RIMD 2210633 (2), *Pseudomonas syringae* pv. tomato strain DC3000 (2), *Enterococcus faecalis* V583, *Pyrococcus horikoshii* OT3, *Bordetella bronchiseptica* RB50, *Streptomyces coelicolor* A3, *Corynebacterium glutamicum* ATCC 13032, *Picrophilus torridus* DSM 9790, *Acinetobacter* sp. ADP1, *Vibrio fischeri* ES114, *Pseudomonas fluorescens* Pf-5, *Sulfolobus acidocaldarius* DSM 639, *Colwellia psychrerythraea* 34H, *Neisseria meningitidis* MC58, *Rhodobacter sphaeroides* 2.4.1, *Vibrio cholerae* O1 biovar eltor strain N16961 (2), *Marinobacter aquaeolei* VT8, *Acinetobacter baumannii* ATCC 17978, *Klebsiella pneumoniae* subsp. pneumoniae MGH 78578, *Streptomyces avermitilis* MA-4680, *Bordetella parapertussis* 12822 (2), *Streptococcus thermophilus* LMG 18311, *Salmonella enterica* subsp. enterica serovar Paratyphi A strain ATCC 9150, and *Anaeromyxobacter dehalogenans* 2CP-C.

Selected cDNAs were cloned into a modified pET vector (Novagen) that fuses a Flag and deca-histidine tag at the C terminus, which are cleavable by TEV protease. Proteins were expressed in *E. coli* BL21(DE3) plysS by a high-throughput format (0.6 ml in a deep-well block) and purified after lysis by sonication using metal affinity purification in a buffer containing *N*-dodecyl-β-D-maltopyranoside. Samples were passed over an analytical size-exclusion column in 12 different detergent-containing mobile phases, which included *N*-dodecyl-β-D-maltopyranoside (DDM), *N*-decyl-β-D-altopyranoside (DM), *N*-nonyl-β-D-altopyranoside (NM), *N*-octyl-β-D-altopyranoside (OM), *N*-octyl-β-D-glucopyranoside (OG), *N*-nonyl-β-D-glucopyranoside (NG) and lauryl dimethyl amine oxide (LDAO). Multi-angle light scattering with refractive index detection was used to analyse the oligomeric state[50]. The *E. coli* and *H. influenzae* proteins were judged to be monodisperse and stable and were passed to scale up.

Scaled-up production and purification. For scale up, in brief, transformed BL21 plysS cells were grown at 37 °C in 2× YT media to an optical density of 0.6–0.8 after being inoculated with 1% of the overnight culture. The culture was induced with 0.4 mM IPTG and continued to grow at 37 °C for another 4 h. The cells were harvested by centrifugation and stored at −80 °C before use. Selenomethionyl (SeMet) TehA was expressed in a similar way, but using SeMet in place of methionine in defined minimal media. Cells were re-suspended in a buffer containing 50 mM Tris-HCl (pH 8.0) and 200 mM NaCl and lysed using a French Press with two passes at 15,000–20,000 p.s.i. Cell debris was removed by centrifugation at 10,000*g* for 20 min, and the membrane fraction was isolated from that supernatant by ultra-centrifugation at 150,000*g* for 1 h.

The membrane fraction was homogenized in a solubilization buffer containing 50 mM Tris (pH 8.0) and 200 mM NaCl, and incubated with a final concentration of 1% (w/v) dodecyl-β-D-maltopyranoside (DDM, Anatrace) for 1 h at 4 °C. The non-dissolved matter was removed by ultracentrifugation at 150,000*g* for 30 min, and the supernatant was loaded to a 5-ml Hitrap Ni^{2+}-NTA affinity column (GE Healthcare), pre-equilibrated with the same solubilization buffer supplemented with 0.05% DDM. After a 20-column-volume buffer wash, the protein was eluted with 250 mM imidazole in the solubilization buffer. The Flag and 10-His tags were removed by adding super TEV at 1:100 mass ratio and incubating at 4 °C overnight. Tag removal was confirmed by SDS–PAGE, and the resulting sample was concentrated to ~10 mg ml^{-1}. Preparative size-exclusion chromatography was carried out on a Superdex-200 column for further purification, removal of TEV protease and the cleaved tag, and for buffer and detergent exchange. The gel-filtration buffer contained 10 mM Tris (pH 8.0), 200 mM NaCl, 1 mM EDTA, 0.5 mM Tris [2-carboxyethyl] phosphine (TCEP), and 2× CMC of detergent. In the case of HiTehA, the protein was well behaved and stable in nearly all tested detergents, and we have purified it from DDM, DM, NM, OM, OG and LDAO.

Protein characterization. We performed N-terminal amino acid sequencing of purified HiTehA and EcTehA before TEV protease treatment. Results from these analyses proved that the true initiating methionine residue is located 14 residues after the one annotated as N terminal. The intervening nucleotide sequence contains a Shine–Delgarno sequence.

For cross-linking experiments, purified HiTehA was incubated with 10 mM disuccinimidyl glutarate at room temperature for 30 min, and 100 mM Tris-HCl pH 7.5 was added to stop the reaction. The incubated sample when run on an 8–25% gradient SDS–PAGE gel showed a ladder consistent with a trimeric structure.

Crystallization and data collection. Purified protein was concentrated to ~10 mg ml^{-1} for initial crystal trials in a Mosquito robot with commercial screens from Hampton research, Emerald Biosystems and Molecular Dimension. We obtained crystals of HiTehA from protein in detergent DDM, DM, NM, OM, OG and LDAO, but only those from LDAO and OG gave diffraction to beyond 4 Å spacings. Crystals that proved useful were all grown at 4 °C using the sitting-drop vapour diffusion method. After extensive optimization, crystals were obtained for diffraction analysis at very high resolution. Wild-type HiTehA and most of its variants were crystallized from 1 mM ZnSO₄, 50 mM HEPES-Na pH 7.8 and 28% PEG600. HiTehA F262A was crystallized from 200 mM Li₂SO₄, 100 mM glycine pH 9.3 and 33% PEG400. Addition of 10 mM spermidine as an additive helped us to obtain slightly better diffracting crystals. Cryoprotection was achieved by adding 5% ethylene glycol or PEG400 to the crystallization solution.

Structure determination and refinement. Native and SeMet single-wavelength anomalous diffraction (SAD) data sets were collected at NSLS beamline X4A and processed using the software HKL2000[51]. Crystals of HiTehA grew in space group *R*3 with $a = b = c = 85.0$ Å and $\alpha = \beta = \gamma = 93.5°$. All subsequent manipulations were done in the hexagonal setting of this space group with $a = b = 96.0$ Å and $c = 136.7$ Å. The asymmetric unit contains one TehA protomer and 65% solvent volume. The structure was determined at 2.0 Å and then extended to 1.5 Å resolution by SAD using selenomethionine-substituted protein crystals. Assessment of data quality for phasing, location of heavy atom sites and initial phases was calculated using the HKL2MAP interface to SHELX programs[52].

All the secondary structure elements were clearly visible in the experimental electron density map. Automatic model building was done in Arp/wArp[53] and completed manually in the program COOT[54]. The model was refined against native data at 1.20 Å resolution using the program Refmac5.5 in CCP4[55], with anisotropic B-factor restrained refinement applied. Subsequent structural analyses of mutant variants were refined as isomorphous structures.

Site-directed mutagenesis. Site-directed mutants were constructed using the QuikChange Site-Directed Mutagenesis Kit (Stratagene) and expressed from pET vectors in *E. coli* BL21(DE3) plysS cells as for the wild-type protein.

Electrophysiology. All constructs were cloned into plasmid pGHME2, linearized and transcribed into cRNA using T7 polymerase (mMessage mMachine, Ambion). Oocytes were injected with 50 nl of cRNA solution each, at a constant concentration of 0.5 mg ml^{-1} for HiTehA and AtSLAC1 constructs, with or without 0.5 mg ml^{-1} of AtOST1 cRNA. Voltage-clamp experiments were performed 2 days after cRNA injection. For mixed expression experiments, 25 nl of cRNA solution was injected for each AtSLAC1 component. Two-microelectrode voltage-clamp recordings were performed to measure HiTehA or AtSLAC1 currents as described[6,7]. The microelectrode solutions contained 3 M KCl. For voltage-clamp current recordings, the bath solution contained 1 mM MgCl₂, 1 mM CaCl₂, 10 mM Mes/Tris (pH 5.6) and 30 mM CsCl; for anion selectivity measurements, the bath solution contained 50 mM Cl⁻, NO₃⁻, or malate, or 30 mM SO₃²⁻ (sodium salts), plus 45 mM Na-gluconate, 1 mM Ca-gluconate₂, 1 mM Mg-gluconate₂, 1 mM K-gluconate, as well as 10 mM Tris/Mes (pH 5.6). Osmolarity was adjusted with D-mannitol to 220 mOsmol kg^{-1}. The bath electrode was a 3 M KCl agar bridge. Voltage-clamp currents were measured in response to 7.5-s-long voltage steps to test potentials that ranged from +50 mV to −110 or −130 mV in 20 mV decrements. Prior to each voltage step the membrane was held at 0 mV for 1.45 s, and following each voltage step the membrane was returned to 0 mV for 2.0 s. *I*–*V* relations for HiTehA or AtSLAC1 channels were generated from currents measured 0.5 s after the start of each test voltage step. The Goldman–Hodgkin–Katz equation was applied to estimate permeability ratios for monovalent ions as described[6]. For divalent anions, the permeability ratios were derived according to ref. 56.

Bioinformatic analysis of SLAC-related proteins. Sequences related to SLAC1 were analysed comprehensively by PSI-BLAST[49]. Searches at $E < 10^{-3}$ starting from five disparate homologues each identified a common pool of over 900 proteins, which when pooled were used for sub-classification into families and subfamilies. Details of these analyses are reported in footnotes to Supplementary Table 1.

Molecular figures. Molecular figures were produced in PyMOL[57].

49. Altschul, S. F. *et al.* Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25**, 3389–3402 (1997).
50. Kendrick, B. S., Kerwin, B. A., Chang, B. S. & Philo, J. S. Online size-exclusion high-performance liquid chromatography light scattering and differential refractometry methods to determine degree of polymer conjugation to proteins

and protein-protein or protein-ligand association states. *Anal. Biochem.* **299,** 136–146 (2001).

51. Otwinowski, Z. & Minor, W. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276,** 307–326 (1997).
52. Pape, P. & Schneider, T. R. HKL2MAP: a graphical user interface for phasing with SHELX programs. *J. Appl. Cryst.* **37,** 843–844 (2004).
53. Perrakis, A., Morris, R. & Lamzin, V. S. Automated protein model building combined with iterative structure refinement. *Nature Struct. Biol.* **6,** 458–463 (1999).
54. Emsley, P. & Cowtan, K. COOT: model-building tools for molecular graphics. *Acta Crystallogr. D* **60,** 2126–2132 (2004).
55. Collaborative Computational Project, Number 4. The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. D* **50,** 760–763 (1994).
56. Fatt, P. & Ginsborg, B. L. The ionic requirements for the production of action potentials in crustacean muscle fibers. *J. Physiol. (Lond.)* **142,** 516–543 (1958).
57. DeLano, W. L. The PyMOL molecular graphics system ⟨http://www.pymol.org⟩ (DeLano Scientific, 2002).

A two-solar-mass neutron star measured using Shapiro delay

P. B. Demorest[1], T. Pennucci[2], S. M. Ransom[1], M. S. E. Roberts[3] & J. W. T. Hessels[4,5]

Neutron stars are composed of the densest form of matter known to exist in our Universe, the composition and properties of which are still theoretically uncertain. Measurements of the masses or radii of these objects can strongly constrain the neutron star matter equation of state and rule out theoretical models of their composition[1,2]. The observed range of neutron star masses, however, has hitherto been too narrow to rule out many predictions of 'exotic' non-nucleonic components[3–6]. The Shapiro delay is a general-relativistic increase in light travel time through the curved space-time near a massive body[7]. For highly inclined (nearly edge-on) binary millisecond radio pulsar systems, this effect allows us to infer the masses of both the neutron star and its binary companion to high precision[8,9]. Here we present radio timing observations of the binary millisecond pulsar J1614-2230[10,11] that show a strong Shapiro delay signature. We calculate the pulsar mass to be $(1.97 \pm 0.04)M_\odot$, which rules out almost all currently proposed[2–5] hyperon or boson condensate equations of state ($M_\odot$, solar mass). Quark matter can support a star this massive only if the quarks are strongly interacting and are therefore not 'free' quarks[12].

In March 2010, we performed a dense set of observations of J1614-2230 with the National Radio Astronomy Observatory Green Bank Telescope (GBT), timed to follow the system through one complete 8.7-d orbit with special attention paid to the orbital conjunction, where the Shapiro delay signal is strongest. These data were taken with the newly built Green Bank Ultimate Pulsar Processing Instrument (GUPPI). GUPPI coherently removes interstellar dispersive smearing from the pulsar signal and integrates the data modulo the current apparent pulse period, producing a set of average pulse profiles, or flux-versus-rotational-phase light curves. From these, we determined pulse times of arrival using standard procedures, with a typical uncertainty of ~1 μs.

We used the measured arrival times to determine key physical parameters of the neutron star and its binary system by fitting them to a comprehensive timing model that accounts for every rotation of the neutron star over the time spanned by the fit. The model predicts at what times pulses should arrive at Earth, taking into account pulsar rotation and spin-down, astrometric terms (sky position and proper motion), binary orbital parameters, time-variable interstellar dispersion and general-relativistic effects such as the Shapiro delay (Table 1). We compared the observed arrival times with the model predictions, and obtained best-fit parameters by χ^2 minimization, using the TEMPO2 software package[13]. We also obtained consistent results using the original TEMPO package. The post-fit residuals, that is, the differences between the observed and the model-predicted pulse arrival times, effectively measure how well the timing model describes the data, and are shown in Fig. 1. We included both a previously recorded long-term data set and our new GUPPI data in a single fit. The long-term data determine model parameters (for example spin-down rate and astrometry) with characteristic timescales longer than a few weeks, whereas the new data best constrain parameters on timescales of the orbital period or less. Additional discussion of the

long-term data set, parameter covariance and dispersion measure variation can be found in Supplementary Information.

As shown in Fig. 1, the Shapiro delay was detected in our data with extremely high significance, and must be included to model the arrival times of the radio pulses correctly. However, estimating parameter values and uncertainties can be difficult owing to the high covariance between many orbital timing model terms[14]. Furthermore, the χ^2 surfaces for the Shapiro-derived companion mass (M_2) and inclination angle (i) are often significantly curved or otherwise non-Gaussian[15]. To obtain robust error estimates, we used a Markov chain Monte Carlo (MCMC) approach to explore the post-fit χ^2 space and derive posterior probability distributions for all timing model parameters (Fig. 2). Our final results for the model

Table 1 | Physical parameters for PSR J1614-2230

| Parameter	Value
Ecliptic longitude (λ)	245.78827556(5)°
Ecliptic latitude (β)	−1.256744(2)°
Proper motion in λ	9.79(7) mas yr^{-1}
Proper motion in β	−30(3) mas yr^{-1}
Parallax	0.5(6) mas
Pulsar spin period	3.1508076534271(6) ms
Period derivative	9.6216(9) × 10^{-21} s s^{-1}
Reference epoch (MJD)	53,600
Dispersion measure*	34.4865 pc cm^{-3}
Orbital period	8.6866194196(2) d
Projected semimajor axis	11.2911975(2) light s
First Laplace parameter ($e\sin\omega$)	1.1(3) × 10^{-7}
Second Laplace parameter ($e\cos\omega$)	−1.29(3) × 10^{-6}
Companion mass	0.500(6)$M_\odot$
Sine of inclination angle	0.999894(5)
Epoch of ascending node (MJD)	52,331.1701098(3)
Span of timing data (MJD)	52,469–55,330
Number of TOAs†	2,206 (454, 1,752)
Root mean squared TOA residual	1.1 μs
Right ascension (J2000)	16 h 14 min 36.5051(5) s
Declination (J2000)	−22° 30′ 31.081(7)′′
Orbital eccentricity (e)	1.30(4) × 10^{-6}
Inclination angle	89.17(2)°
Pulsar mass	1.97(4)$M_\odot$
Dispersion-derived distance‡	1.2 kpc
Parallax distance	>0.9 kpc
Surface magnetic field	1.8 × 10^{8} G
Characteristic age	5.2 Gyr
Spin-down luminosity	1.2 × 10^{34} erg s^{-1}
Average flux density* at 1.4 GHz	1.2 mJy
Spectral index, 1.1–1.9 GHz	−1.9(1)
Rotation measure	−28.0(3) rad m^{-2}

Timing model parameters (top), quantities derived from timing model parameter values (middle) and radio spectral and interstellar medium properties (bottom). Values in parentheses represent the 1σ uncertainty in the final digit, as determined by MCMC error analysis. The fit included both 'long-term' data spanning seven years and new GBT–GUPPI data spanning three months. The new data were observed using an 800-MHz-wide band centred at a radio frequency of 1.5 GHz. The raw profiles were polarization- and flux-calibrated and averaged into 100-MHz, 7.5-min intervals using the PSRCHIVE software package[25], from which pulse times of arrival (TOAs) were determined. MJD, modified Julian date.
* These quantities vary stochastically on $\gtrsim$1-d timescales. Values presented here are the averages for our GUPPI data set.
† Shown in parentheses are separate values for the long-term (first) and new (second) data sets.
‡ Calculated using the NE2001 pulsar distance model[26].

[1]National Radio Astronomy Observatory, 520 Edgemont Road, Charlottesville, Virginia 22093, USA. [2]Astronomy Department, University of Virginia, Charlottesville, Virginia 22094-4325, USA. [3]Eureka Scientific, Inc., Oakland, California 94602, USA. [4]Netherlands Institute for Radio Astronomy (ASTRON), Postbus 2, 7990 AA Dwingeloo, The Netherlands. [5]Astronomical Institute "Anton Pannekoek", University of Amsterdam, 1098 SJ Amsterdam, The Netherlands.
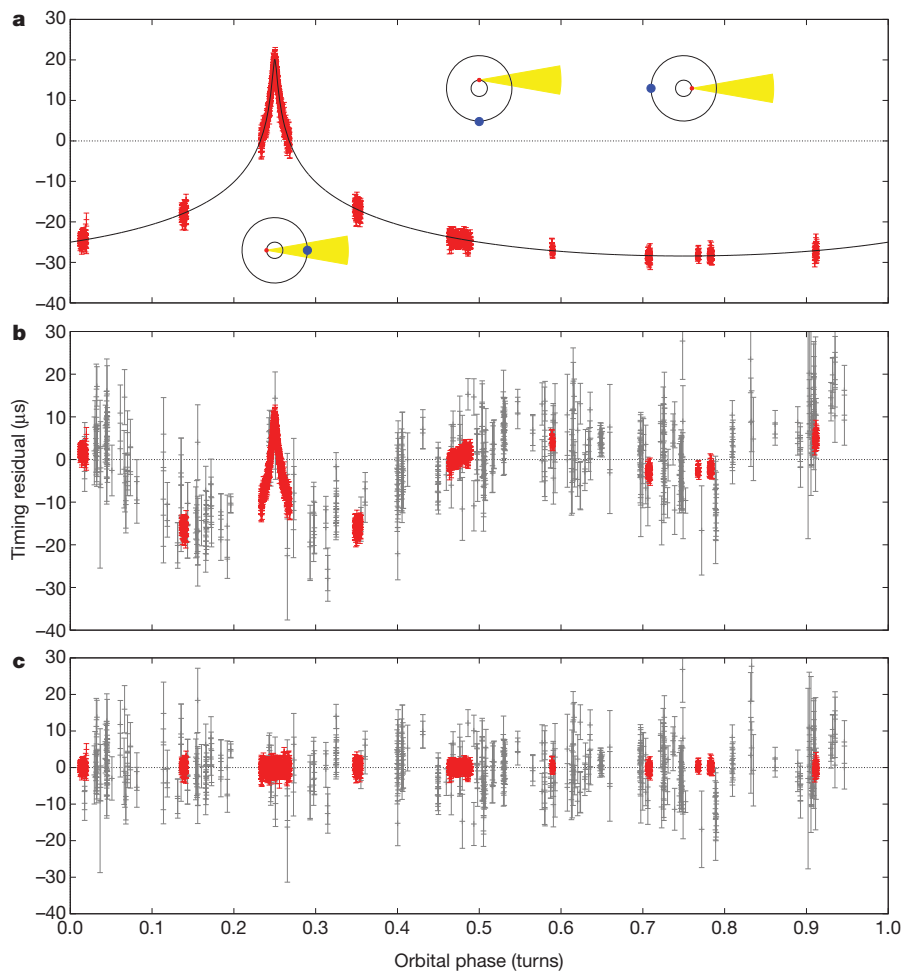
Figure 1 | Shapiro delay measurement for PSR J1614-2230. Timing residual—the excess delay not accounted for by the timing model—as a function of the pulsar's orbital phase. **a**, Full magnitude of the Shapiro delay when all other model parameters are fixed at their best-fit values. The solid line shows the functional form of the Shapiro delay, and the red points are the 1,752 timing measurements in our GBT–GUPPI data set. The diagrams inset in this panel show top-down schematics of the binary system at orbital phases of 0.25, 0.5 and 0.75 turns (from left to right). The neutron star is shown in red, the white dwarf companion in blue and the emitted radio beam, pointing towards Earth, in yellow. At orbital phase of 0.25 turns, the Earth–pulsar line of sight passes nearest to the companion (~240,000 km), producing the sharp peak in pulse delay. We found no evidence for any kind of pulse intensity variations, as from an eclipse, near conjunction. **b**, Best-fit residuals obtained using an orbital model that does not account for general-relativistic effects. In this case, some of the Shapiro delay signal is absorbed by covariant non-relativistic model parameters. That these residuals deviate significantly from a random, Gaussian distribution of zero mean shows that the Shapiro delay must be included to model the pulse arrival times properly, especially at conjunction. In addition to the red GBT–GUPPI points, the 454 grey points show the previous 'long-term' data set. The drastic improvement in data quality is apparent. **c**, Post-fit residuals for the fully relativistic timing model (including Shapiro delay), which have a root mean squared residual of 1.1 μs and a reduced χ^2 value of 1.4 with 2,165 degrees of freedom. Error bars, 1σ.

parameters, with MCMC error estimates, are given in Table 1. Owing to the high significance of this detection, our MCMC procedure and a standard χ^2 fit produce similar uncertainties.

From the detected Shapiro delay, we measure a companion mass of $(0.500 \pm 0.006)M_\odot$, which implies that the companion is a helium–carbon–oxygen white dwarf[16]. The Shapiro delay also shows the binary
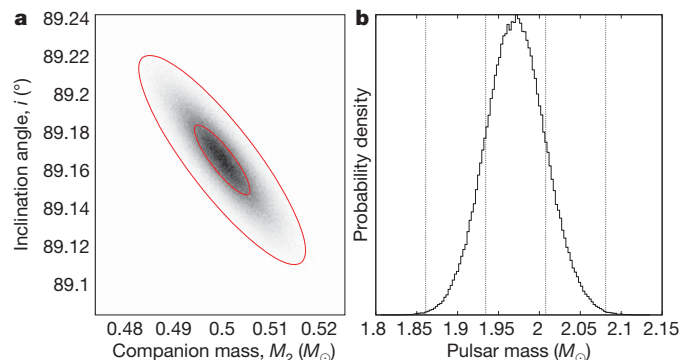

Figure 2 | Results of the MCMC error analysis. a, Grey-scale image shows the two-dimensional posterior probability density function (PDF) in the M_2–i plane, computed from a histogram of MCMC trial values. The ellipses show 1σ and 3σ contours based on a Gaussian approximation to the MCMC results. **b**, PDF for pulsar mass derived from the MCMC trials. The vertical lines show the 1σ and 3σ limits on the pulsar mass. In both cases, the results are very well described by normal distributions owing to the extremely high signal-to-noise ratio of our Shapiro delay detection. Unlike secular orbital effects (for example precession of periastron), the Shapiro delay does not accumulate over time, so the measurement uncertainty scales simply as $T^{-1/2}$, where T is the total observing time. Therefore, we are unlikely to see a significant improvement on these results with currently available telescopes and instrumentation.

system to be remarkably edge-on, with an inclination of $89.17° \pm 0.02°$. This is the most inclined pulsar binary system known at present. The amplitude and sharpness of the Shapiro delay increase rapidly with increasing binary inclination and the overall scaling of the signal is linearly proportional to the mass of the companion star. Thus, the unique combination of the high orbital inclination and massive white dwarf companion in J1614-2230 cause a Shapiro delay amplitude orders of magnitude larger than for most other millisecond pulsars. In addition, the excellent timing precision achievable from the pulsar with the GBT and GUPPI provide a very high signal-to-noise ratio measurement of both Shapiro delay parameters within a single orbit.

The standard Keplerian orbital parameters, combined with the known companion mass and orbital inclination, fully describe the dynamics of a 'clean' binary system—one comprising two stable compact objects—under general relativity and therefore also determine the pulsar's mass. We measure a pulsar mass of $(1.97 \pm 0.04)M_\odot$, which is by far the highest precisely measured neutron star mass determined to date. In contrast with X-ray-based mass/radius measurements[17], the Shapiro delay provides no information about the neutron star's radius. However, unlike the X-ray methods, our result is nearly model independent, as it depends only on general relativity being an adequate description of gravity. In addition, unlike statistical pulsar mass determinations based on measurement of the advance of periastron[18–20], pure Shapiro delay mass measurements involve no assumptions about classical contributions to periastron advance or the distribution of orbital inclinations.

The mass measurement alone of a $1.97M_\odot$ neutron star significantly constrains the nuclear matter equation of state (EOS), as shown in Fig. 3. Any proposed EOS whose mass–radius track does not intersect the J1614-2230 mass line is ruled out by this measurement. The EOSs that produce the lowest maximum masses tend to be those which predict significant softening past a certain central density. This is a
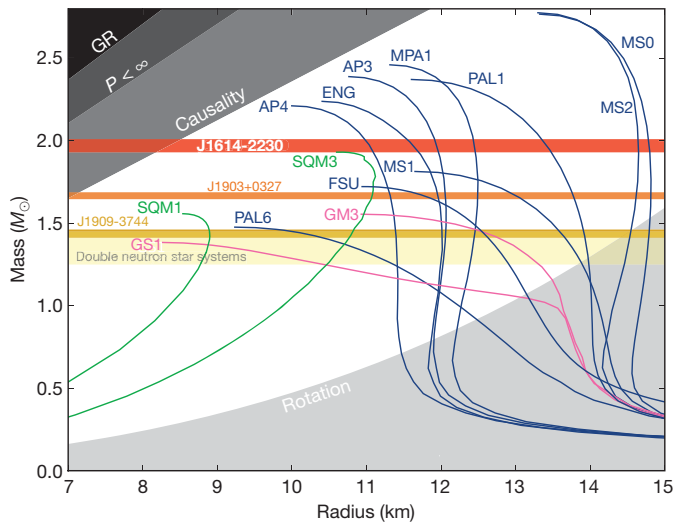
Figure 3 | Neutron star mass–radius diagram. The plot shows non-rotating mass versus physical radius for several typical EOSs[27]: blue, nucleons; pink, nucleons plus exotic matter; green, strange quark matter. The horizontal bands show the observational constraint from our J1614-2230 mass measurement of $(1.97 \pm 0.04)M_\odot$, similar measurements for two other millisecond pulsars[8,28] and the range of observed masses for double neutron star binaries[2]. Any EOS line that does not intersect the J1614-2230 band is ruled out by this measurement. In particular, most EOS curves involving exotic matter, such as kaon condensates or hyperons, tend to predict maximum masses well below $2.0M_\odot$ and are therefore ruled out. Including the effect of neutron star rotation increases the maximum possible mass for each EOS. For a 3.15-ms spin period, this is a $\lesssim 2\%$ correction[29] and does not significantly alter our conclusions. The grey regions show parameter space that is ruled out by other theoretical or observational constraints[2]. GR, general relativity; P, spin period.

common feature of models that include the appearance of 'exotic' hadronic matter such as hyperons[4,5] or kaon condensates[3] at densities of a few times the nuclear saturation density (n_s), for example models GS1 and GM3 in Fig. 3. Almost all such EOSs are ruled out by our results. Our mass measurement does not rule out condensed quark matter as a component of the neutron star interior[6,21], but it strongly constrains quark matter model parameters[12]. For the range of allowed EOS lines presented in Fig. 3, typical values for the physical parameters of J1614-2230 are a central baryon density of between $2n_s$ and $5n_s$ and a radius of between 11 and 15 km, which is only 2–3 times the Schwarzschild radius for a $1.97M_\odot$ star. It has been proposed that the Tolman VII EOS-independent analytic solution of Einstein's equations marks an upper limit on the ultimate density of observable cold matter[22]. If this argument is correct, it follows that our mass measurement sets an upper limit on this maximum density of $(3.74 \pm 0.15) \times 10^{15}\,\mathrm{g\,cm^{-3}}$, or $\sim 10n_s$.

Evolutionary models resulting in companion masses $>0.4M_\odot$ generally predict that the neutron star accretes only a few hundredths of a solar mass of material, and result in a mildly recycled pulsar[23], that is one with a spin period >8 ms. A few models resulting in orbital parameters similar to those of J1614-2230[23,24] predict that the neutron star could accrete up to $0.2M_\odot$, which is still significantly less than the $\gtrsim 0.6M_\odot$ needed to bring a neutron star formed at $1.4M_\odot$ up to the observed mass of J1614-2230. A possible explanation is that some neutron stars are formed massive ($\sim 1.9M_\odot$). Alternatively, the transfer of mass from the companion may be more efficient than current models predict. This suggests that systems with shorter initial orbital periods and lower companion masses—those that produce the vast majority of the fully recycled millisecond pulsar population[23]—may experience even greater amounts of mass transfer. In either case, our mass measurement for J1614-2230 suggests that many other millisecond pulsars may also have masses much greater than $1.4M_\odot$.

1. Lattimer, J. M. & Prakash, M. The physics of neutron stars. *Science* **304**, 536–542 (2004).
2. Lattimer, J. M. & Prakash, M. Neutron star observations: prognosis for equation of state constraints. *Phys. Rep.* **442**, 109–165 (2007).
3. Glendenning, N. K. & Schaffner-Bielich, J. Kaon condensation and dynamical nucleons in neutron stars. *Phys. Rev. Lett.* **81**, 4564–4567 (1998).
4. Lackey, B. D., Nayyar, M. & Owen, B. J. Observational constraints on hyperons in neutron stars. *Phys. Rev. D* **73**, 024021 (2006).
5. Schulze, H., Polls, A., Ramos, A. & Vidaña, I. Maximum mass of neutron stars. *Phys. Rev. C* **73**, 058801 (2006).
6. Kurkela, A., Romatschke, P. & Vuorinen, A. Cold quark matter. *Phys. Rev. D* **81**, 105021 (2010).
7. Shapiro, I. I. Fourth test of general relativity. *Phys. Rev. Lett.* **13**, 789–791 (1964).
8. Jacoby, B. A., Hotan, A., Bailes, M., Ord, S. & Kulkarni, S. R. The mass of a millisecond pulsar. *Astrophys. J.* **629**, L113–L116 (2005).
9. Verbiest, J. P. W. et al. Precision timing of PSR J0437−4715: an accurate pulsar distance, a high pulsar mass, and a limit on the variation of Newton's gravitational constant. *Astrophys. J.* **679**, 675–680 (2008).
10. Hessels, J. et al. in *Binary Radio Pulsars* (eds Rasio, F. A. & Stairs, I. H.) 395 (ASP Conf. Ser. 328, Astronomical Society of the Pacific, 2005).
11. Crawford, F. et al. A survey of 56 midlatitude EGRET error boxes for radio pulsars. *Astrophys. J.* **652**, 1499–1507 (2006).
12. Özel, F., Psaltis, D., Ransom, S., Demorest, P. & Alford, M. The massive pulsar PSR J1614−2230: linking quantum chromodynamics, gamma-ray bursts, and gravitational wave astronomy. *Astrophys. J.* (in the press).
13. Hobbs, G. B., Edwards, R. T. & Manchester, R. N. TEMPO2, a new pulsar-timing package - I. An overview. *Mon. Not. R. Astron. Soc.* **369**, 655–672 (2006).
14. Damour, T. & Deruelle, N. General relativistic celestial mechanics of binary systems. II. The post-Newtonian timing formula. *Ann. Inst. Henri Poincaré Phys. Théor.* **44**, 263–292 (1986).
15. Freire, P. C. C. & Wex, N. The orthometric parameterisation of the Shapiro delay and an improved test of general relativity with binary pulsars. *Mon. Not. R. Astron. Soc.* (in the press).
16. Iben, I. Jr & Tutukov, A. V. On the evolution of close binaries with components of initial mass between 3 solar masses and 12 solar masses. *Astrophys. J Suppl. Ser.* **58**, 661–710 (1985).
17. Özel, F. Soft equations of state for neutron-star matter ruled out by EXO 0748 - 676. *Nature* **441**, 1115–1117 (2006).
18. Ransom, S. M. et al. Twenty-one millisecond pulsars in Terzan 5 using the Green Bank Telescope. *Science* **307**, 892–896 (2005).
19. Freire, P. C. C. et al. Eight new millisecond pulsars in NGC 6440 and NGC 6441. *Astrophys. J.* **675**, 670–682 (2008).
20. Freire, P. C. C., Wolszczan, A., van den Berg, M. & Hessels, J. W. T. A massive neutron star in the globular cluster M5. *Astrophys. J.* **679**, 1433–1442 (2008).
21. Alford, M. et al. Astrophysics: quark matter in compact stars? *Nature* **445**, E7–E8 (2007).
22. Lattimer, J. M. & Prakash, M. Ultimate energy density of observable cold baryonic matter. *Phys. Rev. Lett.* **94**, 111101 (2005).
23. Podsiadlowski, P., Rappaport, S. & Pfahl, E. D. Evolutionary sequences for low- and intermediate-mass X-ray binaries. *Astrophys. J.* **565**, 1107–1133 (2002).
24. Podsiadlowski, P. & Rappaport, S. Cygnus X-2: the descendant of an intermediate-mass X-Ray binary. *Astrophys. J.* **529**, 946–951 (2000).
25. Hotan, A. W., van Straten, W. & Manchester, R. N. PSRCHIVE and PSRFITS: an open approach to radio pulsar data storage and analysis. *Publ. Astron. Soc. Aust.* **21**, 302–309 (2004).
26. Cordes, J. M. & Lazio, T. J. W. NE2001.I. A new model for the Galactic distribution of free electrons and its fluctuations. Preprint at ⟨http://arxiv.org/abs/astro-ph/0207156⟩ (2002).
27. Lattimer, J. M. & Prakash, M. Neutron star structure and the equation of state. *Astrophys. J.* **550**, 426–442 (2001).
28. Champion, D. J. et al. An eccentric binary millisecond pulsar in the Galactic plane. *Science* **320**, 1309–1312 (2008).
29. Berti, E., White, F., Maniopoulou, A. & Bruni, M. Rotating neutron stars: an invariant comparison of approximate and numerical space-time models. *Mon. Not. R. Astron. Soc.* **358**, 923–938 (2005).

LETTER

Detecting excitation and magnetization of individual dopants in a semiconductor

Alexander A. Khajetoorians[1], Bruno Chilian[1], Jens Wiebe[1], Sergej Schuwalow[2], Frank Lechermann[2] & Roland Wiesendanger[1]

An individual magnetic atom doped into a semiconductor is a promising building block for bottom-up spintronic devices and quantum logic gates[1–3]. Moreover, it provides a perfect model system for the atomic-scale investigation of fundamental effects such as magnetism in dilute magnetic semiconductors[4]. However, dopants in semiconductors so far have not been studied by magnetically sensitive techniques with atomic resolution that correlate the atomic structure with the dopant's magnetism. Here we show electrical excitation and read-out of a spin associated with a single magnetic dopant in a semiconductor host. We use spin-resolved scanning tunnelling spectroscopy to measure the spin excitations and the magnetization curve of individual iron surface-dopants embedded within a two-dimensional electron gas confined to an indium antimonide (110) surface. The dopants act like isolated quantum spins the states of which are governed by a substantial magnetic anisotropy that forces the spin to lie in the surface plane. This result is corroborated by our first principles calculations. The demonstrated methodology opens new routes for the investigation of sample systems that are more widely studied in the field of spintronics— that is, Mn in GaAs (ref. 5), magnetic ions in semiconductor quantum dots[3], nitrogen-vacancy centres in diamond[6] and phosphorus spins in silicon[7].

The implementation of future spintronic-based technologies hinges on the control of both spin and charge degrees of freedom of the electrons[8,9]. To this end, manipulating, coupling, and reading individual spins is a prerequisite for solid-state information storage as well as quantum information processing[10,11]. Spins residing in semiconductor matrices provide a viable class of materials owing to their compatibility with conventional fabrication techniques such as molecular beam epitaxy and lithography, which can be used to contact or isolate the spin electronically in two-dimensional electron gases (2DEGs) or quantum dots. Manipulation of spin states in semiconductors, both optically and electrically, has been demonstrated for a variety of systems[2,7]. However, the quantum magnetic behaviour of isolated spins is heavily dependent on the local atomic environment, which affects the magnetic anisotropy energy[12]. A magnetically sensitive atomic-scale method applied to semiconductor systems is thus crucial to understand the correlation between atomic structure and magnetism of single spins in these materials.

The current from the tip of a scanning tunnelling microscope (STM) can excite isolated spins of atoms[12–15] or molecules[16,17] supported on thin decoupling layers and metallic surfaces, permitting the characterization of magnetic anisotropy effects on the individual spin states. On the other hand, a magnetic tip can read out the expectation value of the ground-state spin in a magnetic field—that is, the magnetization curve—of an individual atom on a metallic substrate[18]. Here, we combine these complementary approaches to address a well defined semiconductor model system formed by the spin from an Fe atom that is embedded in a III–V semiconductor 2DEG.

The 2DEG is formed at the (110) surface of n-doped InSb resulting from an accumulation layer induced by the charge transfer of a dilute density of deposited Fe atoms (see Methods, Supplementary Information and Supplementary Fig. 1). The electronic states of such a 2DEG are directly accessible to an STM at subkelvin temperatures[19,20]. It has two occupied sub-bands starting at $E_1 = -80 \pm 20$ meV and $E_2 = -25 \pm 20$ meV below the Fermi energy E_F ($V = 0$ V) (Supplementary Fig. 1). The electrons of these sub-bands have a low effective mass $m^* \approx 0.02 m_e$ (where m_e is the free-electron mass) and an extraordinarily large and negative Landé g-factor $g_{InSb} \approx -45$. The Fe atoms act as localized atomic spins that may weakly couple to the 2DEG[21]. Here, the coverage is low enough (0.1 atoms nm^{-2}) that single Fe atoms behave as isolated entities.

To understand the local structure of this magnetic defect we recorded STM topographs and performed density functional theory (DFT) calculations (Methods). As shown in Fig. 1a, the surface Sb sublattice is imaged as a regular array of protrusions on topographs taken at $V_{stab} < 200$ mV (sample bias)[22]. An Fe atom appears as a slightly asymmetric feature with a circular depression centred in between the rows of the Sb lattice. The relaxed crystal structure is
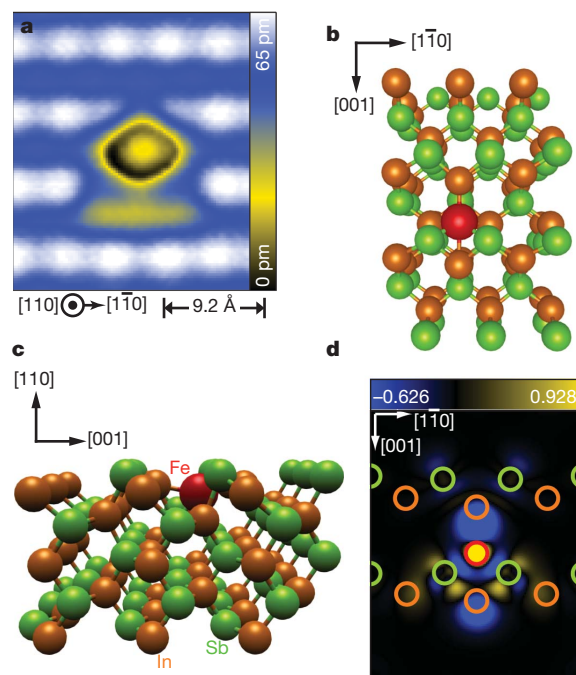
Figure 1 | Fe atoms on InSb(110). a, STM topograph ($I_{stab} = 0.3$ nA, $V_{stab} = 100$ mV) of an Fe atom and the surrounding InSb(110) surface. The Sb sublattice is imaged as protrusions (light blue), while the Fe appears as a protrusion surrounded by a circular depression (yellow) that is centred between the Sb rows. **b** and **c**, Relaxation of the InSb(110) surface with an adsorbed Fe atom as calculated by the DFT method shown in different views. The bulk crystallographic directions are indicated. **d**, DFT-calculated spin density in the (110) surface in units of $10^{-3} e\,\text{Å}^{-3}$.

[1]Institute of Applied Physics, Hamburg University, Jungiusstrasse 11, D-20355 Hamburg, Germany. [2]I. Institute for Theoretical Physics, Hamburg University, Jungiusstrasse 9, D-20355 Hamburg, Germany.

determined from DFT calculations by initially considering the Fe atom to be at the position extracted from the topographs. A top and side view of the relaxed lattice are shown in Fig. 1b and c. While the InSb surface undergoes the well-known relaxation exhibited by the anion (Sb) buckling outward (in the [110] direction) and the cation (In) buckling inward, the Fe core is moved into an interstitial position below both the surface In and Sb atoms.

To detect the spin excitations of the Fe atom, we acquired differential conductance spectra with high energy resolution ($V_{mod} = 40\,\mu V$ root mean square, r.m.s.) in a narrow bias voltage range ($\pm 10\,mV$) around E_F; these spectra are shown in Fig. 2a. Although such spectra measured on the substrate are largely flat and featureless, spectra taken with the same tip above the Fe atom show two distinct steps at both positive and negative sample bias ($V \approx \pm 0.5\,mV$ and $V \approx \pm 1.5\,mV$) that are symmetric to E_F. The intensity of each step amounts to about 25% of the signal at zero bias voltage. These steps are localized above the topographic depression indicating the centre of the Fe atom (Fig. 1a) and vanish at distances greater than one lattice constant. They are thus connected to a local property of the Fe atom core.

Symmetric steps with respect to E_F appearing in conductance spectra are commonly attributed to inelastic tunnelling processes resulting from tunnelling-electron-induced excitations of an adsorbate. For a magnetic atom there are two possibilities: (1) vibrational excitations[23] and (2) spin excitations[13]. Vibronic excitation energies, as estimated by the harmonic potential extracted from the DFT-calculated energy landscape, are at least one order of magnitude larger than the measured step energies. Furthermore, vibronic modes are excited much less efficiently than spin flips and typically comprise only a few per cent of the conductance[23], whereas in the measured spectra every third tunnelling electron induces an excitation (Fig. 2a). Therefore, spin-flip processes must be responsible for the observed steps. In the following, we use a model that links the measured step energies and intensities to the magnitude of the Fe atom's spin and its preferred orientation.
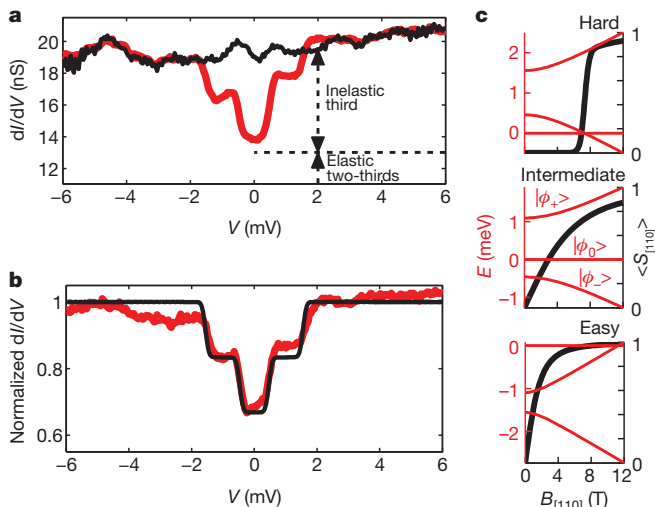
Figure 2 | Inelastic electron tunnelling spectra. a, The spectra ($I_{stab} = 0.16\,nA$, $V_{stab} = 6\,mV$, $V_{mod} = 40\,\mu V$ (r.m.s.), $T = 0.3\,K$) were taken with the tip positioned above the Fe atom (red curve) and above the bare substrate (black curve). The elastic and inelastic contributions are indicated. **b**, Normalized differential conductance (red curve) calculated by dividing the spectrum from the Fe atom by that from the substrate measured with the same tip ($I_{stab} = 0.32\,nA$, $V_{stab} = 10\,mV$, $V_{mod} = 40\,\mu V$ (r.m.s.), $T = 0.3\,K$; data set different from that in **a**). The black curve shows the conductance spectrum calculated from the quantum mechanical model (see text) with $S = 1$ and D and E extracted from the best fit. **c**, Energies of the three eigenstates (red) and expectation value of the [110] component of the spin (black) calculated from the Hamiltonian ($S = 1$, equation (1)) as a function of the strength of a magnetic field in the [110] direction for the three possible cases with [110] being the hard axis, the intermediate axis, and the easy axis, respectively.

For the description of the Fe spin we use the "giant spin" approximation[12,17]:

$$\hat{H} = g_{Fe}\mu_B \mathbf{B}\cdot\hat{\mathbf{S}} + D\hat{S}_z^2 + E\left(\hat{S}_x^2 - \hat{S}_y^2\right) \quad (1)$$

The first term is the Zeeman splitting resulting from an external magnetic field $\mathbf{B}$ with the vector spin operator $\hat{\mathbf{S}} = (\hat{S}_x, \hat{S}_y, \hat{S}_z)$, the Landé g-factor of the Fe atom g_{Fe}, and the Bohr magneton μ_B. The second and the third term describe the so-called magnetic anisotropy energy. It takes into account the spin–orbit interaction in harmonic approximation, and energetically forces the spin to point along certain lattice directions, which are a priori unknown. Diagonalization of the Hamiltonian yields the eigenstates and eigenenergies as a function of $\mathbf{B}$. The steps in the conductance in Fig. 2a appear at bias voltages equal to the energetic separations between the ground state and the excited states at $\mathbf{B} = 0\,T$, which are determined by the parameters D and E. To calculate the step heights that represent the transition probabilities the exchange interaction between the tunnelling electron spin and the atom spin must be considered[24–27]. We use the model from ref. 27, which quantitatively reproduces the spectra measured on magnetic atoms on thin insulating layers[12]. We note that the model implicitly assumes that the transmission through the atom is dominated by one spin character. It will be shown later that this assumption is valid for our case.

The inelastic tunnelling spectra are calculated by assuming different values for the atom spin and for D and E. We find that out of all possible spin values (1/2, 1, 3/2, 2, 5/2) only $S = 1$ properly reproduces both the experimentally observed number of steps as well as the overall intensity of each step as shown by comparison to the normalized spectra in Fig. 2b. To substantiate this result we analysed the DFT calculated charge and spin density around the Fe. Because it acts as a single impurity on the surface, a local-orbital viewpoint remains valid with an additional nearly diagonal orbital density matrix for the d states (Supplementary Information). However, the electronic configuration changes from an atomic Fe ($3d^6 4s^2$) towards ($3d^8 4s^0$) with two unpaired spins that occupy mainly the d_{ζ^2} (ζ parallel to [110]) and the $d_{\xi\upsilon}$ (ξ and υ in (110) plane) orbitals, resulting in a total spin of $S = 1$. As visible in the spin density (Fig. 1d), the spin is strongly concentrated on the Fe atom but partially compensated by the spin distribution in the nearest neighbouring In and Sb atoms. The calculated magnetic moment of the whole unit cell amounts to $2.0\mu_B$. This confirms the experimental result $S = 1$ if we assume $g_{Fe} = 2.0$.

By analysing the step positions in the spectra of several equivalent atoms on different samples, we deduce the following average values for the magnetic anisotropy parameters ($\pm$s.d.): $D = -1.4 \pm 0.3\,meV$ and $E = 0.22 \pm 0.06\,meV$. We note that this choice satisfies the convention to maximize $|D|$ and have $E > 0$, but there are six possible cases corresponding to the orientation of the axes in equation (1) with respect to the three lattice directions ([1$\overline{1}$0], [001] and [110] see Fig. 1). The usual way to distinguish experimentally among these cases is to detect the evolution of the excitation energies as a function of a magnetic field[12]. The eigenenergies of the three eigenstates $|\phi_+\rangle$, $|\phi_-\rangle$ and $|\phi_0\rangle$ (Supplementary Information) in a magnetic field along [110] $B_{[110]}$ for the three distinguishable axis orientations are shown in Fig. 2c. These orientations are: [110] is the most unfavoured 'hard' axis (x), [1$\overline{1}$0] is the 'intermediate' axis (y), and [110] is the favoured 'easy' axis (z). For this particular system, the excitations are strongly masked by the Landau levels emerging in the conductance spectra with the application of a magnetic field (Supplementary Fig. 2), ruling out the use of the usual approach. In the following we will show that the energetic separation of tunnelling electrons with different spin states resulting from spin-splitting of the Landau levels acts as a spin-polarized source that allows us to measure the expectation value of the spin-component along the applied magnetic field (magnetization curve) of the Fe atom. The shape of the magnetization curve indicated in Fig. 2c is strongly sensitive to the orientation of the easy axis, thus allowing us to identify the correct anisotropy case unequivocally.
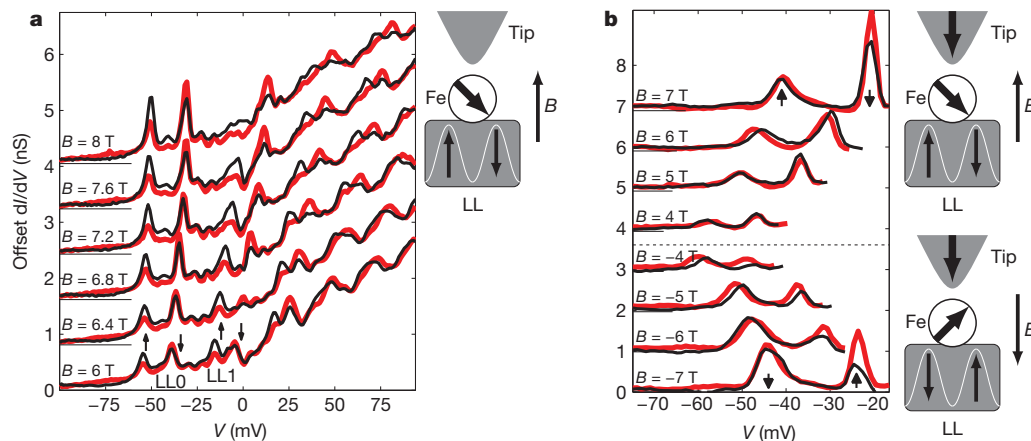
Figure 3 | Spin resolved Landau level spectroscopy. The spectra ($I_{stab} = 0.3$ nA, $V_{stab} = 100$ mV, $V_{mod} = 1$ mV (r.m.s.), $T = 0.3$ K) were taken with the tip positioned above the Fe atom (red curves) and above the bare substrate (black curves) with **B** oriented along the surface normal [110] (curves are offset for clarity as indicated by horizontal zero lines). **a**, Spectra taken with a nonmagnetic tip. Inset, sketch of the directions of the spin in each Landau level (LL) and of the majority electrons in Fe (resulting from the direction of **B**). **b**, Spectra of the lowest spin-split Landau level (LL0) taken with a magnetic tip. Insets, sketches of the directions of the spin in each Landau level and of the majority electrons in Fe and in the foremost tip atom for the two cases $B_{[110]} > 0$ (top) and $B_{[110]} < 0$ (bottom).

Figure 3a shows conductance spectra taken on and off an Fe atom in a wider voltage range as a function of $B_{[110]}$. As expected for a 2DEG, the spectra reveal sharp peaks with energy spacings corresponding to the separation of the Landau levels $heB_{[110]}/m^*$ and the resultant spin-splitting $g_{InSb}\mu_B B_{[110]}$[20]. While both peaks for the lowest Landau level (LL0) corresponding to spin up ($\uparrow$, left) and spin down ($\downarrow$, right) have nearly the same intensity on the substrate, the $\downarrow$ peak has a larger intensity than the $\uparrow$ peak on the Fe atom. A quantitative measure for this asymmetry is obtained by fitting the two peaks to a sum of two Lorentzians with amplitudes a_{right} and a_{left} and defining $A_{LL0} = (a_{right} - a_{left})/(a_{right} + a_{left})$. The dependency of A_{LL0} on $B_{[110]}$ measured at $T = 0.3$ K and $T = 4$ K for different atoms using different non-magnetic tips is shown in Fig. 4a, together with the substrate asymmetry. Although the substrate asymmetry is always below 20%, the atom asymmetry increases with increasing $B_{[110]}$ up to a saturation value of 50%, which is finally achieved for relatively large field

strengths $B_{[110]} > 12$ T. Obviously, the Fe atom that is magnetized by the external magnetic field acts as a spin-filter for tunnelling electrons, as sketched in the inset of Fig. 3a. Owing to the positive g-factor of Fe its majority electron spin is driven into the $\downarrow$ direction, as indicated by the arrow. From the sign of A_{LL0} we see that the $\downarrow$ electrons have a larger transmission than the $\uparrow$ electrons. Because most of the electrons maintain their spin state during tunnelling we can conclude that the transmission through the Fe atom into the tip is dominated by majority electrons. This result validates the assumption (that the transmission through the atom is dominated by one spin character) made by the model we used.

A proof of the interpretation of the observed Landau level asymmetry being due to the spin-filter effect of the atom comes from experiments with spin-polarized tips. Figure 3b shows the conductance spectra of LL0 taken at up- and downwards pointing $B_{[110]}$ using a tip coated with several tens of monolayers of chromium. Such tips are known[18] to act as filters for electrons with a spin component in the axis of the tip ([110]). Consequently, the spectra taken on the substrate already show a considerable asymmetry of $A_{LL0} \approx 30\%$ that directly measures the spin polarization of the tip within the energy range corresponding to LL0. When the direction of $B_{[110]}$ is switched, the substrate asymmetry changes sign, proving that the tip magnetization is not reversed up to $B_{[110]} = \pm 7$ T. The spectra measured on the Fe for upward-pointing $B_{[110]}$ reveal that the atom increases the asymmetry, that is, it acts in the same way as the tip spin filter. In contrast, for $B_{[110]}$ pointing downwards, the atom decreases the asymmetry, even reversing its sign between $B_{[110]} = -6$ T and $B_{[110]} = -7$ T, that is, it counteracts the tip spin filter. The sign change is also visible in the plot of the extracted A_{LL0} from several atoms as a function of $B_{[110]}$ (Fig. 4b). Qualitatively, these experimental findings can be understood by considering the spin orientation of the majority electrons in the tip and in the Fe atom, and of the electrons in the two Landau levels, as given in the inset of Fig. 3b. Quantitatively, the situation is more complex because a considerable fraction of the tunnelling electrons flip their spin as a result of excitations of the atom.

To deduce how A_{LL0} is linked to the atom magnetization, we adapt the model used above to the description of the Landau level asymmetry (Methods). It can be analytically shown that for an unpolarized tip, A_{LL0} is proportional to the component of the atom-spin expectation value along the applied magnetic field (Supplementary Information), that is, A_{LL0} is proportional to the atom magnetization. To predict the measured A_{LL0}, the average values for D and E extracted from the inelastic tunnelling spectra, and a constant tip spin-polarization measured by the

Figure 4 | Landau level asymmetry as a function of the magnetic field. a, The asymmetry A_{LL0} from LL0 measured with a nonmagnetic tip is plotted as a function of $B_{[110]}$. Blue points are taken at $T = 0.3$ K. Red points are taken at $T = 4$ K. Different data sets are indicated by different symbols and were measured with different tips on different samples. The corresponding A_{LL0} from the spectra measured with the same tip on the bare substrate is indicated in grey using the same corresponding symbols. The error bars are calculated from the 95% confidence intervals of the Lorentzian fits to the two Landau level peaks. The solid grey, the solid coloured, and the dashed lines show A_{LL0} calculated from the model for the hard axis case, the intermediate axis case, and the easy axis case, respectively, using D and E determined from the inelastic electron tunnelling spectra (Fig. 2). The shaded area indicates the maximum error of the calculated curve due to the s.d. in D and E. **b**, Same as **a** but from the spectra measured with a magnetic tip. The solid line shows A_{LL0} calculated from the model for the intermediate case assuming a tip spin-polarization of 30%. The shaded area includes the s.d. from the uncertainty in this polarization.

substrate A_{LL0} for the polarized tip are taken into account. The calculated A_{LL0} for the three possible cases of Fig. 2c is plotted in Fig. 4a on top of the measured data. The easy axis and the hard axis cases can be excluded, but a good quantitative agreement is found for the case of the [110] direction being an intermediate axis—that is, the easy axis lies in the (110) plane. For this case the model also correctly predicts the behaviour for the series of atom and tip spin-filters as shown in Fig. 4b. From the experimental values of D and E we can calculate that it costs 0.75 meV and 0.5 meV to rotate the spin from the easy axis to the hard axis and into the intermediate [110] direction, respectively. Using DFT, we calculate that the spin of the Fe atom is preferentially oriented along [1$\bar{1}$0] and the corresponding energies are 1.8 meV and 0.4 meV, in good agreement with the experimental values.

In conclusion, we have shown a marriage of the two complementary methods of spin excitation and magnetization curve measurement for a dopant-associated spin in a semiconductor. The demonstrated methodology not only enables us to measure the magnetic anisotropy, which affects the dopant's spin relaxation time, but also provides a direct means of studying interactions between individual magnetic dopants[28]. Therefore, we anticipate that its application will contribute new microscopic insights into the physics of single and coupled spins in systems that are widely studied in the field of spintronics: Mn in GaAs[4,5], magnetic ions in quantum dots[3], nitrogen-vacancy centres in diamond[6], or phosphorus spins in silicon[7]. Finally, the system studied here is an example of a 'magnetic' 2DEG[29] in which the magnetic dopants might be exchange-coupled to the itinerant 2DEG electrons. An indication for this coupling is given by the hump in the Fe magnetization curve (Fig. 4a) between 6 T and 8 T. Here the Fe magnetization saturates ($A_{LL0} = 50\%$) and then goes back to the value predicted within the isolated-spin model (solid blue line). In the same magnetic field range, the local 2DEG magnetization oscillates, as proved by the consecutive E_F crossing of the $\downarrow$ and $\uparrow$ Landau levels (LL1) in Fig. 3a. Future experiments with larger dopant density will probably show interesting effects of magnetism in this type of diluted magnetic semiconductor[21].

METHODS SUMMARY

The ultrahigh-vacuum STM and the tungsten tip preparation are described elsewhere[19]. We selected tips exhibiting minimal tip-induced band-bending by avoiding tips showing tip-induced quantum-dot states or unreasonable band gaps[20]. Magnetic tips were prepared by coating them with several tens of monolayers of chromium[18,19]. Commercial n-doped InSb single crystals of three different dopings (carrier concentrations at 77 K: 4×10^{15} cm^{-3}, 6.5×10^{15} cm^{-3} and 2×10^{16} cm^{-3}) were cleaved under ultrahigh-vacuum[20]. Fe is deposited onto the cold surface ($T < 25$ K), resulting in a coverage of $(1 \pm 0.5) \times 10^{13}$ atoms cm^{-2}. Measurements were done on atoms far away from any defects or neighbouring Fe (Supplementary Fig. 1). STM topographs were recorded in constant-current mode at a stabilization current I_{stab} with a stabilization voltage V_{stab} applied to the sample. dI/d$V(V)$ curves are taken via lock-in technique with open feedback and a modulation voltage V_{mod} ($f = 828$ Hz) added to V. Details of the DFT calculations are given in the Supplementary Information.

To predict A_{LL0} the model from ref. 27 is modified. The spin-split Landau levels below/above E_F give access to the relative tunnelling probability for each tunnelling electron initial/final spin state. Because V is above all excitation energies, all channels (elastic and inelastic) have to be considered:

$$\text{spin-}m\text{-LL intensity} \propto \sum_{n,n',m'} P_T(|\phi_n\rangle) \left| \sum_{M_T} A_{M_T,n,m} A_{M_T,n',m'} \right|^2 f_{tip}(m') \quad (2)$$

Here, $P_T(|\phi_n\rangle) = \exp(-E_n/k_BT)/\Sigma_j \exp(-E_j/k_BT)$ is the occupation probability for the initial atom state $|\phi_n\rangle$ at temperature T, $f_{tip}(m')$ is the fraction of spin-m' density-of-states at the tip apex determined from the substrate A_{LL0}, and $A_{M_T,n,m} = \langle\phi_n; 1/2, m|S_T = S+1/2, M_T\rangle$ is the overlap of the intermediate total spin state with the asymptotic product state composed of the atom and tunnelling electron states. The indices $m, m' \in \{+1/2, -1/2\}$ label the tunnelling electron initial and final states and $n, n' \in \{+1, 0, -1\}$ label the initial and final atom states. Using equation (2), A_{LL0} can be calculated as a function of **B**. For non-magnetic tips at 300 mK, A_{LL0} is proportional to the Fe magnetization and saturates at 0.5 (Supplementary Information).

1. Tang, J.-M., Levy, J. & Flatté, M. E. All-electrical control of single ion spins in a semiconductor. *Phys. Rev. Lett.* **97,** 106803 (2006).
2. Hanson, R. & Awschalom, D. D. Coherent manipulation of single spins in semiconductors. *Nature* **453,** 1043–1049 (2008).
3. Le Gall, C. *et al.* Optical spin orientation of a single manganese atom in a semiconductor quantum dot using quasiresonant photoexcitation. *Phys. Rev. Lett.* **102,** 127402 (2009).
4. Kitchen, D., Richardella, A., Tang, J.-M., Flatté, M. E. & Yazdani, A. Atom-by-atom substitution of Mn in GaAs and visualization of their hole-mediated interactions. *Nature* **442,** 436–439 (2006).
5. Yakunin, A. M. *et al.* Warping a single Mn acceptor wavefunction by straining the GaAs host. *Nature Mater.* **6,** 512–515 (2007).
6. Neumann, P. *et al.* Quantum register based on coupled electron spins in a room-temperature solid. *Nature Phys.* **6,** 249–253 (2010).
7. Fuechsle, M. *et al.* Spectroscopy of few-electron single-crystal silicon quantum dots. *Nature Nanotechnol.* **5,** 502–505 (2010).
8. Žutić, I., Fabian, J. & Das Sarma, S. Spintronics: fundamentals and applications. *Rev. Mod. Phys.* **76,** 323–410 (2004).
9. Awschalom, D. D. & Flatté, M. E. Challenges for semiconductor spintronics. *Nature Phys.* **3,** 153–159 (2007).
10. Kane, B. E. A silicon-based nuclear spin quantum computer. *Nature* **393,** 133–137 (1998).
11. Loss, D. & DiVincenzo, D. P. Quantum computation with quantum dots. *Phys. Rev. A* **57,** 120–126 (1998).
12. Hirjibehedin, C. F. *et al.* Large magnetic anisotropy of a single atomic spin embedded in a surface molecular network. *Science* **317,** 1199–1203 (2007).
13. Heinrich, A. J., Gupta, J. A., Lutz, C. P. & Eigler, D. M. Single-atom spin-flip spectroscopy. *Science* **306,** 466–469 (2004).
14. Loth, S. *et al.* Controlling the state of quantum spins with electric currents. *Nature Phys.* **6,** 340–344 (2010).
15. Balashov, T. *et al.* Magnetic anisotropy and magnetization dynamics of individual atoms and clusters of Fe and Co on Pt(111). *Phys. Rev. Lett.* **102,** 257203 (2009).
16. Chen, X. *et al.* Probing superexchange interaction in molecular magnets by spin-flip spectroscopy and microscopy. *Phys. Rev. Lett.* **101,** 197208 (2008).
17. Tsukahara, N. *et al.* Adsorption-induced switching of magnetic anisotropy in a single iron(II) phthalocyanine molecule on an oxidized Cu(110) surface. *Phys. Rev. Lett.* **102,** 167203 (2009).
18. Meier, F., Zhou, L., Wiebe, J. & Wiesendanger, R. Revealing magnetic interactions from single-atom magnetization curves. *Science* **320,** 82–86 (2008).
19. Wiebe, J. *et al.* A 300 mK ultra-high vacuum scanning tunneling microscope for spin-resolved spectroscopy at high energy resolution. *Rev. Sci. Instrum.* **75,** 4871–4879 (2004).
20. Hashimoto, K. *et al.* Quantum Hall transition in real space: from localized to extended states. *Phys. Rev. Lett.* **101,** 256802 (2008).
21. Mochizuki, T., Masutomi, R. & Okamoto, T. Evidence for two-dimensional spin-glass ordering in submonolayer Fe films on cleaved InAs surfaces. *Phys. Rev. Lett.* **101,** 267204 (2008).
22. Whitman, L. J., Stroscio, J. A., Dragoset, R. A. & Celotta, R. J. Scanning-tunneling-microscopy study of InSb(110). *Phys. Rev. B* **42,** 7288–7291 (1990).
23. Stipe, B. C., Rezaei, M. A. & Ho, W. Single-molecule vibrational spectroscopy and microscopy. *Science* **280,** 1732–1735 (1998).
24. Fernández-Rossier, J. Theory of single-spin inelastic tunneling spectroscopy. *Phys. Rev. Lett.* **102,** 256802 (2009).
25. Fransson, J. Spin inelastic electron tunneling spectroscopy on local spin adsorbed on surface. *Nano Lett.* **9,** 2414–2417 (2009).
26. Persson, M. Theory of inelastic electron tunneling from a localized spin in the impulsive approximation. *Phys. Rev. Lett.* **103,** 050801 (2009).
27. Lorente, N. & Gauyacq, J.-P. Efficient spin transitions in inelastic electron tunneling spectroscopy. *Phys. Rev. Lett.* **103,** 176601 (2009).
28. Zhou, L. *et al.* Strength and directionality of surface Ruderman-Kittel-Kasuya-Yosida interaction mapped on the atomic scale. *Nature Phys.* **6,** 187–191 (2010).
29. Harris, J. G. E. *et al.* Magnetization measurements of magnetic two-dimensional electron gases. *Phys. Rev. Lett.* **86,** 4644 (2001).

Author Contributions A.A.K. and B.C. performed the experiments. A.A.K., B.C. and J.W. did the data analysis. B.C. did the modelling. S.S. did the DFT calculations. J.W., A.A.K. and B.C. wrote the paper. All authors discussed the results and commented on the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to J.W. (jwiebe@physnet.uni-hamburg.de).

LETTER

The evolution of the marine phosphate reservoir

Noah J. Planavsky[1,2], Olivier J. Rouxel[2,3], Andrey Bekker[4], Stefan V. Lalonde[5], Kurt O. Konhauser[5], Christopher T. Reinhard[1] & Timothy W. Lyons[1]

Phosphorus is a biolimiting nutrient that has an important role in regulating the burial of organic matter and the redox state of the ocean–atmosphere system[1]. The ratio of phosphorus to iron in iron-oxide-rich sedimentary rocks can be used to track dissolved phosphate concentrations if the dissolved silica concentration of sea water is estimated[2–5]. Here we present iron and phosphorus concentration ratios from distal hydrothermal sediments and iron formations through time to study the evolution of the marine phosphate reservoir. The data suggest that phosphate concentrations have been relatively constant over the Phanerozoic eon, the past 542 million years (Myr) of Earth's history. In contrast, phosphate concentrations seem to have been elevated in Precambrian oceans. Specifically, there is a peak in phosphorus-to-iron ratios in Neoproterozoic iron formations dating from ~750 to ~635 Myr ago, indicating unusually high dissolved phosphate concentrations in the aftermath of widespread, low-latitude 'snowball Earth' glaciations. An enhanced postglacial phosphate flux would have caused high rates of primary productivity and organic carbon burial and a transition to more oxidizing conditions in the ocean and atmosphere. The snowball Earth glaciations and Neoproterozoic oxidation are both suggested as triggers for the evolution and radiation of metazoans[6,7]. We propose that these two factors are intimately linked; a glacially induced nutrient surplus could have led to an increase in atmospheric oxygen, paving the way for the rise of metazoan life.

In almost all modern aquatic systems, primary production of organic matter is typically thought to be limited by either phosphorus or bioavailable nitrogen[8]. Temporally extended deficiencies in fixed nitrogen availability are buffered by biological fixation of a virtually limitless supply of atmospheric N_2. By contrast, phosphorus is sourced primarily by weathering of continental materials; accordingly, it is generally thought that phosphorus ultimately limits net primary productivity on geological timescales[8,9]. An estimate of marine phosphate reservoir size through time is therefore essential to unravel basic aspects of biological and geochemical evolution[10].

Ratios of phosphorus to iron in ferric oxides scale with ambient concentrations of dissolved phosphate ($[P_D]$), as predicted by distribution coefficient (K_D) relationships: $[P_D] = (1/K_D)P/Fe$ (ref. 4). P/Fe ratios in ferric oxyhydroxides within hydrothermal plumes emanating from mid-ocean ridges remain constant during transport[3]. Similarly, P/Fe ratios in modern iron-oxide-rich sediments seem to remain essentially constant or show only slight decreases during burial, despite mineralogical transformations[5]. Consequently, P/Fe ratios in ferruginous sediments can be used to track dissolved phosphate concentrations in ancient sea water[2]. Because the K_D value for phosphate-iron oxyhydroxide sorption varies inversely with dissolved silica concentrations owing to competitive adsorption of aqueous silica species[11], it is also important to consider the evolution of the silica cycle when using P/Fe ratios as a palaeoproxy. Marine silica concentrations have varied drastically through Earth's history, significantly influencing phosphate sorption by iron oxyhydroxides.

The data for this study (~700 individual samples of iron-oxide-rich rocks) include new results and those obtained from a comprehensive literature survey. Consistent with previous studies in which iron formations were used to decipher ancient seawater chemistry[2,12], we passed our samples through a series of strict filters to select for authigenic iron-rich rocks that most probably retain bulk seawater signatures. All samples have a negligible detrital component and contain only minor amounts of pyrite, siderite, and manganese phases (Supplementary Information).

We identify four well-defined stages in P/Fe ratios in iron-oxide-rich rocks through time (Fig. 1). These stages reflect both shifts in the size of the marine phosphate reservoir and the evolution of the global silica cycle. Stages one and two, defined by distal hydrothermal sediments from fourteen different localities of Phanerozoic age (<542 Myr), span the Quaternary period to the Cretaceous and the Jurassic period to the Cambrian, respectively. The molar P/Fe ratios multiplied by one hundred ($P/Fe_{(100)}$) in stage one yield an average of 2.55 with a range of <1 to 8.6 and a standard deviation of 1.2. In stage two, there is an average $P/Fe_{(100)}$ ratio of 0.38 with a range from <1 to 1.8 and a standard deviation of 0.26 (Fig. 1).

The marked change in P/Fe ratios between stages one and two is coincident with the initial radiation of diatoms, when marine silica concentrations are thought to have decreased substantially[13,14]. Dissolved marine silica concentrations are assumed to have been <0.1 mM since the Cretaceous and, taking a conservative estimate, ~0.67 mM between the Cambrian and the mid-Jurassic, which is near cristobalite saturation (see Supplementary Information for a discussion of constraints on dissolved silica concentrations). Recent
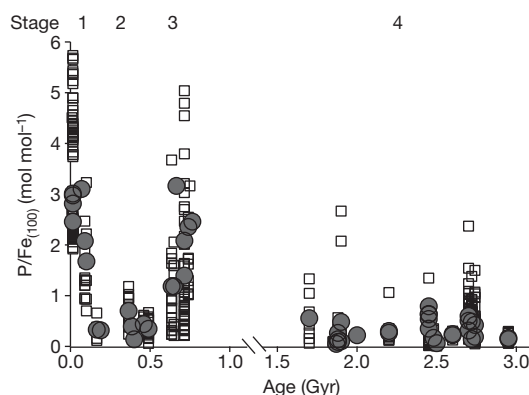
Figure 1 | P/Fe molar ratios through time in iron-oxide-rich distal hydrothermal sediments and iron formations with low amounts of siliciclastic input. Open squares are individual samples; filled circles are formation averages. The P/Fe ratio reflects the size of the marine phosphate reservoir; phosphate sorption onto ferric oxyhydroxides follows a distribution coefficient (K_D) relationship. The ratio is also influenced by the concentration of dissolved silica, because phosphate and silica hydroxides compete for sorption sites on ferric oxyhydroxides. Two outliers are not shown ($P/Fe_{(100)} = 8.6$ 90 Myr ago and $P/Fe_{(100)} = 6.8$ 750 Myr ago). See Supplementary Information for a box plot of the data.

[1]Department of Earth Sciences, University of California, Riverside, California 92521, USA. [2]Department of Marine Chemistry and Geochemistry, Woods Hole Oceanographic Institution, Woods Hole, Massachusetts 02543, USA. [3]Université Européenne de Bretagne, European Institute for Marine Studies, Technopôle Brest-Iroise, Place Nicolas Copernic, 29280 Plouzané, France. [4]Department of Geological Sciences, University of Manitoba, Winnipeg, Manitoba R3T 2N2, Canada. [5]Department of Earth and Atmospheric Sciences, University of Alberta, Edmonton, Alberta T6G 2E3, Canada.

experimental results indicate that an approximately sevenfold increase in dissolved silica from modern concentrations (that is, from <0.1 to 0.67 mM) would cause an 85.6% decrease in the amount of phosphate sorbed to ferric oxides (Supplementary Fig 1). This decrease is virtually identical to the magnitude of the observed increase in P/Fe ratios occurring subsequent to the expansion of siliceous phytoplankton (85%), when dissolved silica concentrations would have decreased. Thus, when viewed in light of varying silica content, marine phosphate concentrations seem to have been roughly constant though the Phanerozoic, in accordance with independent estimates for marine phosphate concentrations[15].

Stage three occurred during the Cryogenian period (~750–620 Myr ago). Samples for this time interval are from seven iron formations associated with the low-latitude, snowball Earth glaciations[6] and contain an average P/Fe$_{(100)}$ ratio of 1.96 with a range from <1 to 6.8 and a standard deviation of 1.2. This average is markedly higher than those seen during the early and middle Phanerozoic. Dissolved marine silica concentrations in the Neoproterozoic era were probably high relative to the Phanerozoic; the radiation of radiolarians and siliceous sponges in the earliest Phanerozoic resulted in a shift to a biologically controlled silica cycle and probably caused a decrease in marine silica concentrations[13]. Therefore, Cryogenian iron formations point to very high marine phosphate concentrations. Using even the most conservative estimates for dissolved silica concentrations, assuming that concentrations were similar to those in the early Phanerozoic (~0.67 mM), P/Fe ratios in Cryogenian iron formations suggest that marine dissolved phosphate concentrations were more than five times greater than Phanerozoic levels. Neoproterozoic iron formations were deposited in a shelf or slope setting under shallower conditions than the majority of deep-water (substorm wave base to abyssal depths) deposits in our compilation. Because of the non-conservative, nutrient-type behaviour of dissolved phosphorus in the ocean, shallow waters would be expected to show signs of phosphate depletion, making the anomalous enrichments we see in the shallow Cryogenian iron formation samples even more remarkable. The occurrence of high P/Fe ratios in seven separate, geographically widespread successions (see Supplementary Information for formation details) supports our assertion that the observed high P/Fe ratios reflect global conditions during glacial periods in the Cryogenian rather than conditions unique to isolated basins.

Stage four spans the Palaeoproterozoic era and the Archaean eon (1.7–3.0 billion years ago) and is represented by iron formations and distal hydrothermal sediments from 24 localities. The P/Fe$_{(100)}$ ratio in stage four is 0.37 with a range from <1 to 2.9 and a standard deviation of 0.42. This average P/Fe$_{(100)}$ ratio is approximately equal to that found in early- and mid-Phanerozoic rocks but is significantly less than ratios found in iron formations deposited during the snowball Earth glacial period. Dissolved silica concentrations in the Archaean and Palaeoproterozoic oceans may have been as low as cristobalite saturation (~0.67 mM) but it is more likely that they were near amorphous silica saturation[14,16] (~2.2 mM). Correspondingly, phosphate concentrations in Earth's early oceans are estimated to have been, at minimum, equivalent to Phanerozoic levels but are more likely to have been several times higher (~4 times assuming 2.2 mM dissolved silica). However, because most of the Archaean and early Proterozoic samples in our compilation contain mixed-valence iron oxides, some caution should be exercised when making comparisons to exclusively ferric-iron-dominated rocks (Supplementary Information).

Changes in the global biogeochemical cycle of phosphorus can be related to the evolution of Earth's surface conditions (Fig. 2). It is likely that the major removal fluxes for phosphate from modern oceans were attenuated during the Precambrian. Ferric oxyhydroxides represent a substantial sink in modern oceans[17], but the importance of this sink would have been lower in the Precambrian because of less phosphate sorption onto ferric oxides at high concentrations of dissolved silica. In addition, substantial portions of the deep ocean were probably anoxic
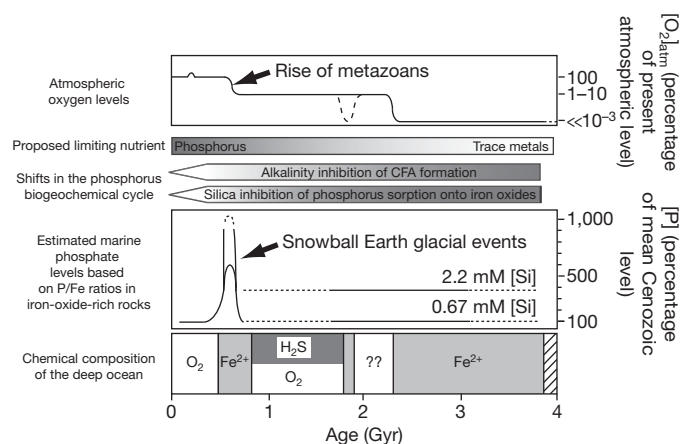
Figure 2 | Model for the coevolution of atmospheric and oceanic redox state and limiting nutrients for marine primary productivity. The redox model is from refs 10, 30. Phosphate concentrations are extrapolated from average P/Fe ratios for individual formations. Our compilation of P/Fe data suggests that there were elevated seawater phosphate concentrations in the Precambrian and a peak in phosphate levels associated with the Neoproterozoic snowball Earth glaciations. This late Precambrian increase in dissolved phosphorus concentration may have stimulated high rates of organic carbon burial and a corresponding increase in atmospheric oxygen levels—paving the way for the rise of metazoans. CFA, carbonate fluorapatite. Square brackets denote concentration.

before and even during the Neoproterozoic[18], which would have removed the large phosphate sink associated with ferric oxyhydroxide formation during off-axis hydrothermal alteration of basalts[17]. Perhaps more importantly, the formation of carbonate fluorapatite during early diagenesis, the largest marine phosphate sink today[19], was probably much less effective during the early Precambrian. Carbonate fluorapatite solubility scales with carbonate alkalinity[20], which was almost certainly high before the onset of enzymatic carbonate formation in the late Neoproterozoic[21]. Lastly, high marine phosphate concentrations during the extensive Cryogenian glaciations are expected, given that weathering rates in modern glaciated settings are higher than those in comparable unglaciated catchments[22]. Enhanced post- and syn-glacial phosphorus delivery to marine systems results in part from an elevated detrital flux to, and high dissolution rates within, proglacial environments[22]. Importantly, in the Neoproterozoic, before soil stabilization by vascular plants, the temporal extent of enhanced phosphorus delivery from glaciated catchments was probably much greater than in the Pleistocene.

Because phosphorus is believed to be the nutrient ultimately controlling marine primary productivity on geological timescales[9], elevated marine phosphate concentrations should lead to higher levels of organic matter production and increased carbon burial. However, because of high biological metal demands, especially in diazotrophic (N$_2$-fixing) organisms, trace elements may also limit primary productivity. Non-ferrous trace-element stress is likely to have been severe in Earth's early oceans. Under an essentially anoxic atmosphere during the Archaean, there would have been limited continental weathering and delivery of dissolved redox-sensitive metals to the oceans[23] (for example cobalt, cadmium, molybdenum and vanadium, which are common cofactors and are crucial in many major metabolic processes, including nitrogen assimilation and fixation). Our study points to high phosphate concentrations in the Archaean and Palaeoproterozoic oceans, thereby strengthening earlier arguments asserting that non-ferrous trace metals, rather than phosphorus, were the most important factors limiting carbon fixation in Earth's early biosphere. Under a later oxidizing atmosphere and widespread euxinic (anoxic and sulphidic) or oxic conditions in the ocean, trace-element stress is also possible[24]. For instance, trace metals (for example, iron) seem to limit nitrogen fixation in regions of the modern ocean[24,25].

We suggest that a combination of upwelling iron-rich waters[18] and significantly elevated marine phosphorus concentrations following the snowball Earth glaciations would have caused a nutrient surplus—stimulating high rates of primary productivity and increased organic carbon burial. Unprecedented continental phosphorus fluxes would be expected following these glaciations (during post-glacial and inter-glacial time periods), given the extraordinary extent and duration of Cryogenian ice cover and the high levels of phosphorus delivery expected from glaciated catchments. Persistently high carbonate carbon isotope values for significant time periods of the Cryogenian[26] confirm this increase in organic carbon burial. Additionally, perturbations to the carbon cycle connected to the snowball Earth events, for instance extensive methane clathrate release, may have muted the carbonate carbon isotope signature for high post-glacial organic carbon burial[6,27]. A long-lived, glacially induced nutrient surplus and a corresponding organic carbon burial event would have resulted in a shift to more oxidizing ocean–atmosphere conditions in the late Neoproterozoic, because net burial of organic carbon results in a corresponding rise in atmospheric O_2 (ref. 10). The evolution and ecological expansion of metazoans is largely dependent on the oxidation state of marine systems[7]. Therefore, this redox shift could have paved the way for the rise of metazoans—providing a mechanistic explanation for the intimate link[6,28,29] between the snowball Earth events and early animal evolution.

METHODS SUMMARY

Data in the compilation reflect our analytical efforts and a literature survey and include distal hydrothermal sediments and samples from iron formations (Supplementary Table 1). The criteria used to filter the data are outlined in Supplementary Information. The new trace and major element concentrations were determined using a ThermoFinnigan Element II inductively coupled plasma mass spectrometer operated at Woods Hole Oceanographic Institution following a three-acid digest. Analytical precision and accuracy for our measurements were checked by multiple analyses of the geostandards IF-G and BHVO-1, and reproducibility was better than 5%. Reproducibility of literature data is estimated to be better than 10%. See Supplementary Information for additional method details.

1. Holland, H. D. *The Chemical Evolution of the Atmosphere and Oceans* 598 (Princeton Univ. Press, 1984).
2. Bjerrum, C. J. & Canfield, D. E. Ocean productivity before about 1.9 Gyr ago limited by phosphorus adsorption onto iron oxides. *Nature* **417,** 159–162 (2002).
3. Edmonds, H. N. & German, C. R. Particle geochemistry in the Rainbow hydrothermal plume, Mid-Atlantic Ridge. *Geochim. Cosmochim. Acta* **68,** 759–772 (2004).
4. Feely, R. A., Trefry, J. H., Lebon, G. T. & German, C. R. The relationship between P/Fe and V/Fe ratios in hydrothermal precipitates and dissolved phosphate in seawater. *Geophys. Res. Lett.* **25,** 2253–2256 (1998).
5. Poulton, S. W. & Canfield, D. E. Co-diagenesis of iron and phosphorus in hydrothermal sediments from the southern East Pacific Rise: implications for the evaluation of paleoseawater phosphate concentrations. *Geochim. Cosmochim. Acta* **70,** 5883–5898 (2006).
6. Hoffman, P. F. & Schrag, D. P. The snowball Earth hypothesis: testing the limits of global change. *Terra Nova* **14,** 129–155 (2002).
7. Knoll, A. H. & Carroll, S. B. Early animal evolution: emerging views from comparative biology and geology. *Science* **284,** 2129–2137 (1999).
8. Howarth, R. W. Nutrient limitation of net primary production in marine ecosystems. *Annu. Rev. Ecol. Syst.* **19,** 89–110 (1988).
9. Tyrrell, T. The relative influences of nitrogen and phosphorus on oceanic primary production. *Nature* **400,** 525–531 (1999).
10. Holland, H. D. The oxygenation of the atmosphere and oceans. *Phil. Trans. R. Soc. B* **361,** 903–915 (2006).
11. Konhauser, K. O., Lalonde, S. V., Amskold, L. & Holland, H. D. Was there really an Archean phosphate crisis? *Science* **315,** 1234 (2007).
12. Konhauser, K. O. *et al.* Oceanic nickel depletion and a methanogen famine before the Great Oxidation Event. *Nature* **458,** 750–753 (2009).
13. Racki, G. & Cordey, F. Radiolarian palaeoecology and radiolarites: is the present the key to the past? *Earth Sci. Rev.* **52,** 83–120 (2000).
14. Siever, R. The silica cycle in the Precambrian. *Geochim. Cosmochim. Acta* **56,** 3265–3272 (1992).
15. Arvidson, R. S., Mackenzie, F. T. & Guidry, M. W. MAGic: a Phanerozoic model for the geochemical cycling of major rock-forming components. *Am. J. Sci.* **306,** 135–190 (2006).
16. Maliva, R. G., Knoll, A. H. & Simonson, B. M. Secular change in the Precambrian silica cycle: insights from chert petrology. *Geol. Soc. Am. Bull.* **117,** 835–845 (2005).
17. Wheat, C. G., Feely, R. A. & Mottl, M. J. Phosphate removal by oceanic hydrothermal processes: an update of the phosphorus budget in the oceans. *Geochim. Cosmochim. Acta* **60,** 3593–3608 (1996).
18. Canfield, D. E. *et al.* Ferruginous conditions dominated later Neoproterozoic deep-water chemistry. *Science* **321,** 949–952 (2008).
19. Ruttenberg, K. C. & Berner, R. A. Authigenic apatite formation and burial in sediments from non-upwelling, continental-margin environments. *Geochim. Cosmochim. Acta* **57,** 991–1007 (1993).
20. Jahnke, R. A. The synthesis and solubility of carbonate fluorapatite. *Am. J. Sci.* **284,** 58–78 (1984).
21. Ridgwell, A. & Zeebe, R. E. The role of the global carbonate cycle in the regulation and evolution of the Earth system. *Earth Planet. Sci. Lett.* **234,** 299–315 (2005).
22. Föllmi, K. B., Hosein, R., Arn, K. & Steinmann, P. Weathering and the mobility of phosphorus in the catchments and forefields of the Rhône and Oberaar glaciers, central Switzerland: implications for the global phosphorus cycle on glacial–interglacial timescales. *Geochim. Cosmochim. Acta* **73,** 2252–2282 (2009).
23. Scott, C. *et al.* Tracing the stepwise oxygenation of the Proterozoic ocean. *Nature* **452,** 457–460 (2008).
24. Martin, J. Glacial-interglacial CO₂ change: the iron hypothesis. *Paleoceanography* **5,** 1–13 (1990).
25. Wu, J. F., Sunda, W., Boyle, E. A. & Karl, D. M. Phosphate depletion in the western North Atlantic Ocean. *Science* **289,** 759–762 (2000).
26. Halverson, G. P., Hoffman, P. F., Schrag, D. P., Maloof, A. C. & Rice, H. N. Toward a Neoproterozoic composite carbon-isotope record. *Geol. Soc. Am. Bull.* **117,** 1181–1207 (2005).
27. Jiang, G., Kennedy, M. J. & Christie-Blick, N. Stable isotope evidence for methane seeps in Neoproterozoic postglacial cap carbonates. *Nature* **426,** 822–826 (2003).
28. Cohen, P. A., Knoll, A. H. & Kodner, R. B. Large spinose microfossils in Ediacaran rocks as resting stages of early animals. *Proc. Natl Acad. Sci. USA* **106,** 6519–6524 (2009).
29. Love, G. D. *et al.* Fossil steroids record the appearance of Demospongiae during the Cryogenian period. *Nature* **457,** 718–721 (2008).
30. Canfield, D. E. The early history of atmospheric oxygen: homage to Robert A. Garrels. *Annu. Rev. Earth Planet. Sci.* **33,** 1–36 (2005).

Author Contributions All authors were involved in the writing and design of this study. A.B. and N.J.P. collected samples for this study, and N.J.P. and O.J.R. analysed them. N.J.P. and S.L. compiled literature data.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to T.W.L (timothy.lyons@ucr.edu).

LETTER

Deformation of the lowermost mantle from seismic anisotropy

Andy Nowacki[1], James Wookey[1] & J-Michael Kendall[1]

The lowermost part of the Earth's mantle—known as D″—shows significant seismic anisotropy, the variation of seismic wave speed with direction[1-5]. This is probably due to deformation-induced alignment of $MgSiO_3$-post-perovskite (ppv), which is believed to be the main mineral phase present in the region. If this is the case, then previous measurements of D″ anisotropy, which are generally made in one direction only, are insufficient to distinguish candidate mechanisms of slip in ppv because the mineral is orthorhombic. Here we measure anisotropy in D″ beneath North and Central America, where material from subducting oceanic slabs impinges[6] on the core–mantle boundary, using shallow as well as deep earthquakes to increase the azimuthal coverage in D″. We make more than 700 individual measurements of shear wave splitting in D″ in three regions from two different azimuths in each case. We show that the previously assumed[2,3,7] case of vertical transverse isotropy (where wave speed shows no azimuthal variation) is not possible, and that more complicated mechanisms must be involved. We test the fit of different $MgSiO_3$-ppv deformation mechanisms to our results and find that shear on (001) is most consistent with observations and the expected shear above the core–mantle boundary beneath subduction zones. With new models of mantle flow, or improved experimental determination of the dominant ppv slip

systems, this method will allow us to map deformation at the core–mantle boundary and link processes in D″, such as plume initiation, to the rest of the mantle.

Studies of D″ anisotropy in the Caribbean are numerous[2-4,7-9] because of an abundance of deep earthquakes in South America and seismometers in North America, and show approximately 1% shear wave anisotropy. These studies mostly compare the horizontally polarized (SH) and vertically polarized (SV) shear waves, assuming vertical transverse isotropy, a kind of anisotropy in which the shear wave velocity V_S varies only with the angle away from the vertical. With this assumption, SH leads SV here, corresponding to $\varphi' = +90°$ in our notation (Fig. 1c). A further limitation is their use of only one azimuth of rays in D″: this cannot distinguish vertical transverse isotropy from the case of an arbitrarily tilted axis of rotational symmetry in which wave speed does not vary when the axis dips towards the receivers or stations (tilted transverse isotropy). An improvement on this situation can be made by using crossing ray paths in D″ (ref. 10), but this relies on having the correct source–receiver geometry, which is not possible beneath North America using only deep earthquakes. We address this issue beneath the Caribbean by incorporating measurements from shallow earthquakes in our data set, and thus reduce the symmetry of the anisotropy which must be assumed.
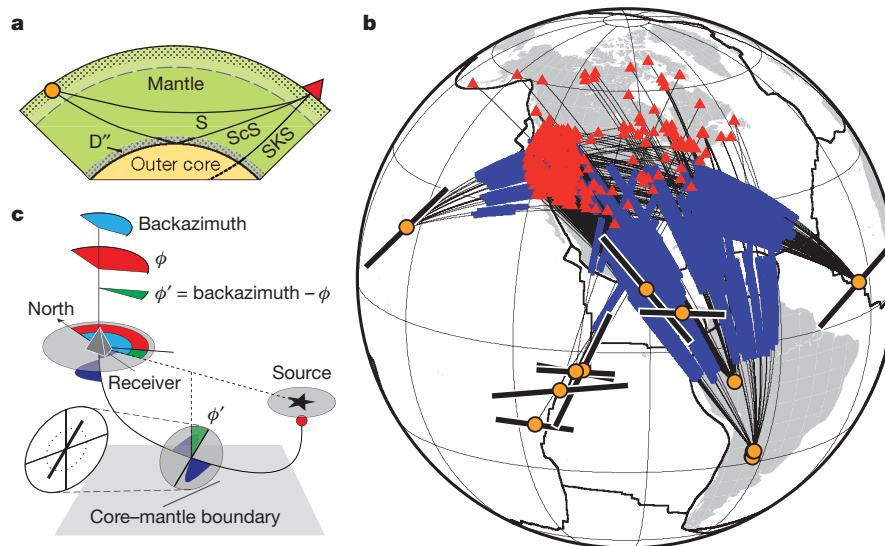
Figure 1 | Source–receiver geometry, and explanation of φ'. a, Earth section with ray paths for the S, ScS and SKS phases. The stippled upper mantle and grey D″ are anisotropic. S turns above D″; ScS samples it. **b,** Shown are seismic stations (triangles), earthquake epicentres (yellow circles), ray paths (thin black lines) and ray paths in a 250-km-thick D″ (blue lines). The measured source-side shear-wave splitting parameters for shallow earthquakes are shown as black bars beneath circles (bar length corresponds to delay time, orientation represents fast direction, largest delay time is 2.4 s). We note that fast orientations of shear-wave splitting in the upper mantle beneath shallow earthquakes on plate boundaries are either generally very closely parallel to the plate-spreading direction (the East Pacific Rise and the Mid-Atlantic Ridge), or to the subduction zone trench (Central America). **c,** Relationship of the measured fast directions in the geographic (φ) and ray (φ') reference frames. Because the ScS phase is nearly horizontal for most of its travel through D″, we define $\varphi' = $ backazimuth $- \varphi$, which corresponds to the polarization away from the vertical of the fast shear wave. In terms of transverse anisotropy, $\varphi' = \pm 90°$ is compatible with vertical transverse isotropy, and $-90° < \varphi' < 90°$ implies tilted transverse isotropy. This can also be thought of as the plane normal to the rotational symmetry axis being tilted from the horizontal, or dipping, at $(90 - \varphi')°$.

[1]Department of Earth Sciences, University of Bristol, Wills Memorial Building, Queen's Road, Bristol, BS8 1RJ, UK.

We measure anisotropy in D″ using differential splitting in S and ScS (respectively direct and reflected from the core–mantle boundary) phases using an approach described by refs 10 and 11. Both phases travel through the same region of the upper mantle, but only ScS samples D″ (Fig. 1a). Given that the majority of the lower mantle is relatively isotropic[12], by removing the splitting introduced in the upper mantle we can measure the splitting that occurs only in D″ (see Supplementary Information). Earthquakes in South and Central America, Hawaii, the East Pacific Rise and the Mid-Atlantic Ridge, detected at North American stations, provide a dense coverage of crossing rays that traverse D″ beneath southern North America and the Caribbean (Fig. 1b). Three distinct regions are covered (Fig. 2), each sampled along two distinct azimuths. The Caribbean (region 'S') has previously been well studied[1,4,8], but the northeast ('E') and southwest ('W') USA have not.

Stacked results along each azimuth in the three regions give splitting parameters shown in Fig. 2 and listed in Supplementary Table 3. We discuss results in terms of the delay time (δt) and ray frame fast orientation (φ'; Fig. 1c). The primary observation is that D″ everywhere shows anisotropy of between 0.8% and 1.5% (assuming a uniform 250-km-thick D″ layer). Along south–north (region 'S') and southeast–northwest (region 'E') ray paths, from deep South American events (approximately 200 measurements), $\delta t = (1.45 \pm 0.55)$ s, implying shear wave anisotropy of about 0.8%. Fast orientations are approximately parallel to the core–mantle boundary ($\varphi' \approx 90°$). This agrees with previous studies made along similar azimuths[4,7–9], including the presence of some small variation in φ' of up to $\pm 15°$ (refs 4 and 8). Such variations could be approximated as vertical transverse isotropy over the region. Detailed results are shown in Supplementary Figs 1 and 11. Notably, however, oblique to the approximately south–north raypaths in the Caribbean, fast directions are at least 40° away from parallel to the core–mantle boundary (region 'S': $\delta t = 1.68$ s, $\varphi' \approx -42°$; region 'E': $\delta t = 1.28$ s, $\varphi' \approx 45°$). In region 'W', both azimuths show φ' about 10°–15° from the horizontal in D″, with $\delta t \approx 1.2$ s. Hence, nowhere are our measurements compatible with vertical transverse isotropy, because we do not find $\varphi' = \pm 90°$ within error in both directions for any region.
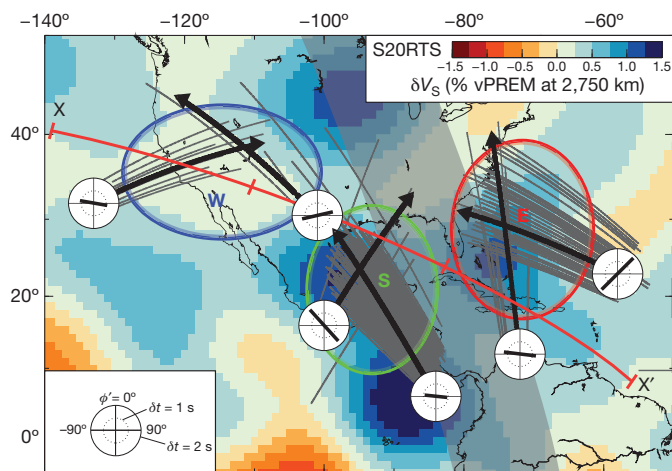
Figure 2 | Multi-azimuth stacked shear wave splitting results in each region. Shown are individual D″ ray paths of ScS phases used in stacks (thin grey lines); representative mean ray paths in D″ of stacked measurements (thick black lines, arrows indicate direction of travel); plots of splitting parameters for each stack at the start of the path (white circles with black bars, angle indicates φ', length indicates δt). The colour shading beneath is the variation of V_S at 2,750 km depth (about 150 km above the core–mantle boundary) in the S20RTS model[30] compared to the Preliminary Reference Earth Model (PREM). The thick red line X–X′ is the cross-section shown in Fig. 3a. The shaded region shows the approximate strike of the Farallon plate predicted at 2,500 km (ref. 6). The three study regions ('W', 'S' and 'E') are indicated by circled areas. Supplementary Fig. 2 shows the approximate finite-frequency zone of sensitivity for ScS in D″.

A likely mechanism for the production of anisotropy in D″ is the lattice-preferred orientation (LPO) of anisotropic mineral phases present above the core–mantle boundary such as (Mg,Fe)O, MgSiO$_3$-perovskite and MgSiO$_3$-ppv. These may give rise to kinds of anisotropy more complicated than tilted transverse isotropy with lower symmetries, which are compatible with our two-azimuth measurements. We investigate the possibility of LPO in ppv leading to the observed anisotropy rather than other phases because of its probable abundance in seismically fast regions of the lowermost mantle beneath North America and its relatively large anisotropy. (Mg,Fe)O and perovskite seem poor candidates for D″ anisotropy because (Mg,Fe)O is equally abundant in the lower mantle above D″, which appears to be relatively isotropic[12], and perovskite is the dominant phase there. Although (Mg,Fe)O may be strongly anisotropic and mechanically weaker than ppv[13–15], and therefore might take up more deformation and align more fully, ppv is also highly anisotropic and is the most abundant phase, meaning a lower degree of alignment of ppv can produce just as much anisotropy as more alignment of (Mg,Fe)O. Therefore LPO in ppv is our preferred mineralogical mechanism.

Different candidate mechanisms for LPO development in ppv from deformation by dislocation creep have been proposed: slip systems of $[\bar{1}10](110)$ (refs 16–18) and $[100](010)$ (refs 19–21) have been inferred from experimental and theoretical methods. Recent experimental work[22] has also suggested that the $[100](001)$ system may be plausible, which is appealing because it appears to best-match the first-order anisotropic signature of the lowermost mantle[23–26].

Our results can differentiate between these candidate mechanisms if we assume that most of the measured anisotropy in D″ is a result of deformation-induced LPO in ppv, and we have an accurate estimate of the mantle flow where we measure anisotropy. At present, such models of mantle deformation are in their infancy, but we can nonetheless make inferences from broad-scale trends in subduction and global V_S models. We calculate the orientations of the shear planes and slip directions that are compatible with our measurements for the three slip systems in ppv. Aggregate elastic constants for the $[\bar{1}10](110)$ and $[001](010)$ systems are taken from deformation experiments[17,20]; we use single-crystal elastic constants from first-principles calculations[23,25] for the $[100](001)$ system. These planes and directions are plotted in Fig. 3. We also produce the shear planes predicted for cases of perovskite and MgO (Supplementary Fig. 11).

At present, there is some disagreement in detail between different *ab initio* elastic constants for ppv[23,27]. We use those of ref. 23 for consistency with experimental studies. Another source of uncertainty may be the extrapolation of results of deformation experiments[16,17,20,22] to lowermost mantle conditions.

To guide our interpretation of the results, we can appeal to the broadly analogous situation of finite strain and olivine LPO associated with passive upwelling beneath a mid-ocean ridge. Models indicate that, near the centre of the upwelling, directions of maximum finite extension dip away from the centre, and become more horizontal with distance from the ridge[28]. Corresponding features beneath downwellings are found in convection models of the lower mantle—inclined deformation dipping towards the downwelling centre[29]. Regions 'E' and 'S' are either side of the apparent centre of the downwelling Farallon slab[6,30] (Figs 2, 3), which strikes roughly northwest–southeast, so we postulate northeast–southwest slip directions on inclined shear planes with an opposite sense of dip (that is, dipping southwest for region 'E', northeast for region 'S'). Further away from the downwelling, in region 'W', more horizontal flow is expected and hence a horizontal shear plane with northeast–southwest slip directions.

All three considered slip systems have orientations that can explain the data, but the predictions of the $[100](001)$ slip system (Fig. 3) best-match the above criteria. The $[\bar{1}10](110)$ system is arguably the least plausible, because it requires complex flow further from the downwelling (region 'W') where a simpler horizontal flow pattern is expected. We cannot yet completely rule out the $[100](010)$ system; more
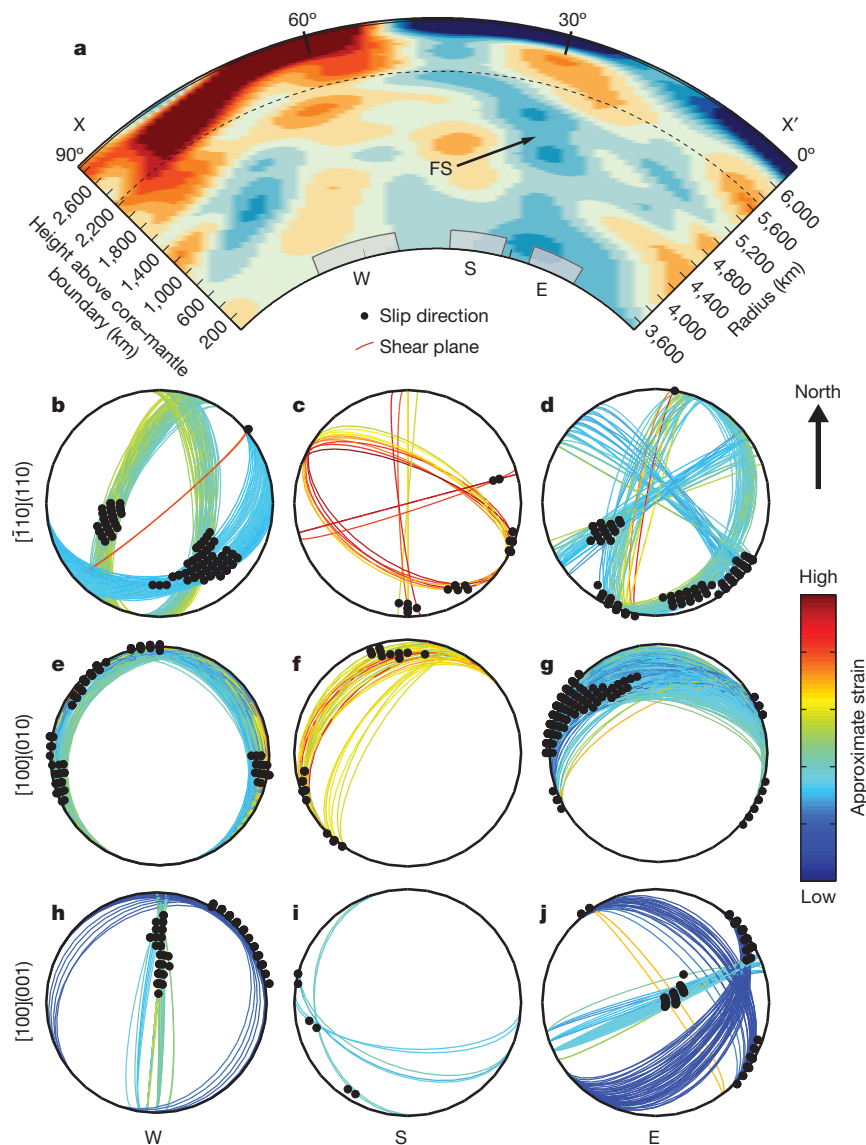
Figure 3 | Section through study region and compatible shear planes for candidate ppv slip systems. a, Cross-section through V_S model S20RTS traversing the study region, as indicated in Fig. 2. The approximate regions 'W', 'S' and 'E' in D″ are drawn. Colours indicate V_S as for Fig. 2. The inferred location of the Farallon slab from high V_S is labelled 'FS'. **b–j**, Orientations of potential elastic models that are compatible with the observed anisotropy in D″. Shown are upper-hemisphere equal-area projections looking down the Earth radial direction (vertical) of the possible shear planes (coloured lines) and slip directions (black circles) in ppv for each slip system. The colour of the shear planes indicates the amount of strain required to produce them according to the arbitrary colour scale, right. The three slip mechanisms [$\bar{1}$10](110) (**b–d**), [100](010) (**e–g**) and [100](001) (**h–i**) are tested in each region (left to right, 'W', 'S', 'E'). Up is north. There are usually two sets of planes, because two azimuths of measurements are not sufficient to define the planes uniquely in the orthorhombic symmetry of the models.

rigorous flow modelling in the region is required to resolve this issue conclusively.

The slip systems predicted for perovskite and MgO (Supplementary Fig. 12) seem less likely, particularly where the measured splitting is high. The presence of perovskite versus ppv in D″ in region 'S', for instance, cannot account for the high anisotropy inferred, and shear planes and directions for MgO are mostly very steep.

D″ anisotropy might also arise from shape-preferred orientation of seismically distinct material over sub-wavelength scales. This would lead to tilted transverse isotropy behaviour[2], with which our observations are compatible. In this case, we can interpret our results simply by finding the common plane, normal to the rotational symmetry axis, from the two azimuths and φ'. These planes are shown in Supplementary Fig. 2.

In each region, the tilted transverse isotropy plane dips approximately in the same way as for the [100](010) case, that is, southwest, southeast and south in regions 'W', 'S' and 'E' respectively, by between

26–52° (Supplementary Fig. 2). However, there is no constraint on the slip direction, and especially in regions 'S' and 'E', where the dip is about 50°, it is hard to correlate the transverse isotropy planes with a candidate plane of deformation based on V_S, and models of deformation suggest that strain in such slab-parallel orientations is unlikely. For this reason and because the post-perovskite phase explains other D″ properties[25], we favour the mineralogical interpretation at present, in which all tested ppv mechanisms are in some agreement with our results, and the [100](001) slip system in ppv is most compatible with our observations.

We have made significant progress towards using D″ anisotropy to measure deformation in the lowermost mantle. Assuming that anisotropy in D″ is caused by the alignment of ppv, we may suggest which slip system dominates LPO, though without more detailed models of mantle flow there is still doubt as to the likely orientation of slip planes and directions in the lowermost mantle. As more reliable estimates of

the type of deformation we expect in well-studied regions become available, and as numerical and physical experiments further indicate the mechanisms by which the material in D″ deforms, our observations of seismic anisotropy may become very useful in the mapping of dynamic processes at the core–mantle boundary.

METHODS SUMMARY

We measured differential shear wave splitting between S and ScS (reflected from the core–mantle boundary) recorded at about 500 seismic stations in North and Central America, using events of moment magnitude $M_{\mathrm{w}} \geq 5.7$ and epicentral distance $55°–82°$ (Supplementary Table 3). Data were bandpass-filtered between 0.001 Hz and 0.3 Hz to remove noise. We analysed splitting in the phases using the minimum eigenvalue technique (Supplementary Fig. 3). We correct for upper mantle anisotropy using published[31,32] SKS (seismic waves travelling as shear waves in the mantle, compression waves in the outer core) splitting measurements at stations showing little variation of parameters with backazimuth—corresponding to simple upper mantle anisotropy—where there are measurements along similar backazimuths to S–ScS used here. Measuring splitting in S with a receiver-side correction gives an estimate of the source-side splitting beneath the earthquake (Fig. 1b; Supplementary Table 3). Both corrections are applied when analysing ScS: the measurement is thus of splitting in D″ alone.

We confirm that the only source of splitting in our measurements is D″ by comparing: (1) splitting in S from a deep event with that in SKS; (2) the source-side anisotropy with SKS measurements at the source; (3) the initial polarization of S after analysis with that predicted by the GlobalCMT moment tensor solution (http://www.globalcmt.org/); (4) the consistency of measurements when correcting with real SKS and randomized receiver corrections; (5) φ' and δt along the same ray paths for deep and shallow events, correcting the latter for upper mantle anisotropy. (See online-only Methods and Supplementary Figs 5–9 for details.)

Orientations of shear planes and slip directions in each slip system of ppv are computed by grid search over the elastic constants[16,20,25], which are rotated about the three principal axes. Shear wave splitting is calculated, and orientations which are compatible with the observations are plotted. The constants are scaled linearly away from the isotropic case to fit the observations, and this scaling is shown by colour (Fig. 3b–j), qualitatively representing strain.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 21 January; accepted 6 September 2010.

1. Kendall, J.-M. & Silver, P. Constraints from seismic anisotropy on the nature of the lowermost mantle. *Nature* **381,** 409–412 (1996).
2. Kendall, J.-M. & Silver, P. G. in *The Core–Mantle Boundary Region* (eds Gurnis, M., Wysession, M. E., Knittle, E. & Buffett, B. A.) 97–118 (American Geophysical Union, 1998).
3. Lay, T., Williams, Q., Garnero, E. J., Kellogg, L. & Wysession, M. E. in *The Core–Mantle Boundary Region* (eds Gurnis, M., Wysession, M. E., Knittle, E. & Buffett, B. A.) 299–318 (American Geophysical Union, 1998).
4. Maupin, V., Garnero, E. J., Lay, T. & Fouch, M. J. Azimuthal anisotropy in the D″ layer beneath the Caribbean. *J. Geophys. Res. Solid Earth* **110,** B08301 (2005).
5. Long, M. D. Complex anisotropy in D″ beneath the eastern Pacific from SKS–SKKS splitting discrepancies. *Earth Planet. Sci. Lett.* **283,** 181–189 (2009).
6. Ren, Y., Stutzman, E., van der Hilst, R. D. & Besse, J. Understanding seismic heterogeneities in the lower mantle beneath the Americas from seismic tomography and plate tectonic history. *J. Geophys. Res. Solid Earth* **112,** B01302 (2007).
7. Kendall, J.-M. & Nangini, C. Lateral variations in D″ below the Caribbean. *Geophys. Res. Lett.* **23,** 399–402 (1996).
8. Garnero, E. J., Maupin, V., Lay, T. & Fouch, M. J. Variable azimuthal anisotropy in Earth's lowermost mantle. *Science* **306,** 259–261 (2004).
9. Rokosky, J. M., Lay, T. & Garnero, E. J. Small-scale lateral variations in azimuthally anisotropic D″ structure beneath the Cocos plate. *Earth Planet. Sci. Lett.* **248,** 411–425 (2006).
10. Wookey, J. & Kendall, J.-M. Constraints on lowermost mantle mineralogy and fabric beneath Siberia from seismic anisotropy. *Earth Planet. Sci. Lett.* **275,** 32–42 (2008).
11. Wookey, J., Kendall, J.-M. & Rumpker, G. Lowermost mantle anisotropy beneath the north Pacific from differential S–ScS splitting. *Geophys. J. Int.* **161,** 829–838 (2005).
12. Meade, C., Silver, P. & Kaneshima, S. Laboratory and seismological observations of lower mantle isotropy. *Geophys. Res. Lett.* **22,** 1293–1296 (1995).
13. Karki, B., Wentzcovitch, R., de Gironcoli, S. & Baroni, S. First-principles determination of elastic anisotropy and wave velocities of MgO at lower mantle conditions. *Science* **286,** 1705–1707 (1999).
14. Long, M. D., Xiao, X., Jiang, Z., Evans, B. & Karato, S. Lattice preferred orientation in deformed polycrystalline (Mg,Fe)O and implications for seismic anisotropy in D″. *Phys. Earth Planet. Inter.* **156,** 75–88 (2006).
15. Yamazaki, D. & Karato, S. Fabric development in (Mg,Fe)O during large strain, shear deformation: implications for seismic anisotropy in Earth's lower mantle. *Phys. Earth Planet. Inter.* **131,** 251–267 (2002).
16. Merkel, S. *et al.* Deformation of (Mg,Fe)SiO₃ post-perovskite and D″ anisotropy. *Science* **316,** 1729–1732 (2007).
17. Merkel, S. *et al.* Plastic deformation of MgGeO₃ post-perovskite at lower mantle pressures. *Science* **311,** 644–646 (2006).
18. Oganov, A., Martonak, R., Laio, A., Raiteri, P. & Parrinello, M. Anisotropy of Earth's D″ layer and stacking faults in the MgSiO₃ post-perovskite phase. *Nature* **438,** 1142–1144 (2005).
19. Carrez, P., Ferré, D. & Cordier, P. Implications for plastic flow in the deep mantle from modelling dislocations in MgSiO₃ minerals. *Nature* **446,** 68–70 (2007).
20. Yamazaki, D., Yoshino, T., Ohfuji, H., Ando, J. & Yoneda, A. Origin of seismic anisotropy in the D″ layer inferred from shear deformation experiments on post-perovskite phase. *Earth Planet. Sci. Lett.* **252,** 372–378 (2006).
21. Iitaka, T., Hirose, K., Kawamura, K. & Murakami, M. The elasticity of the MgSiO₃ post-perovskite phase in the Earth's lowermost mantle. *Nature* **430,** 442–445 (2004).
22. Okada, T., Yagi, T., Niwa, K. & Kikegawa, T. Lattice-preferred orientations in post-perovskite-type MgGeO₃ formed by transformations from different pre-phases. *Phys. Earth Planet. Inter.* **180,** 195–202 (2010).
23. Stackhouse, S., Brodholt, J. P., Wookey, J., Kendall, J.-M. & Price, G. D. The effect of temperature on the seismic anisotropy of the perovskite and post-perovskite polymorphs of MgSiO₃. *Earth Planet. Sci. Lett.* **230,** 1–10 (2005).
24. Tsuchiya, T., Tsuchiya, J., Umemoto, K. & Wentzcovitch, R. Phase transition in MgSiO₃ perovskite in the Earth's lower mantle. *Earth Planet. Sci. Lett.* **224,** 241–248 (2004).
25. Wookey, J., Stackhouse, S., Kendall, J.-M., Brodholt, J. P. & Price, G. D. Efficacy of the post-perovskite phase as an explanation for lowermost-mantle seismic properties. *Nature* **438,** 1004–1007 (2005).
26. Wookey, J. & Kendall, J.-M. in *Post-Perovksite: The Last Mantle Phase Transition* (eds Hirose, K., Brodholt, J., Lay, T. & Yuen, D.) 171–189 (American Geophysical Union, 2007).
27. Wentzcovitch, R., Tsuchiya, T. & Tsuchiya, J. MgSiO₃ postperovskite at D″ conditions. *Proc. Natl Acad. Sci. USA* **103,** 543–546 (2006).
28. Blackman, D. *et al.* Teleseismic imaging of subaxial flow at mid-ocean ridges: traveltime effects of anisotropic mineral texture in the mantle. *Geophys. J. Int.* **127,** 415–426 (1996).
29. McNamara, A., van Keken, P. & Karato, S. Development of finite strain in the convecting lower mantle and its implications for seismic anisotropy. *J. Geophys. Res. Solid Earth* **108,** 2230 (2003).
30. Ritsema, J., van Heijst, H. J. & Woodhouse, J. H. Complex shear wave velocity structure imaged beneath Africa and Iceland. *Science* **286,** 1925–1928 (1999).
31. Evans, M., Kendall, J.-M. & Willemann, R. Automated SKS splitting and upper-mantle anisotropy beneath Canadian seismic stations. *Geophys. J. Int.* **165,** 931–942 (2006).
32. Wuestefeld, A., Bokelmann, G., Barruol, G. & Montagner, J.-P. Identifying global seismic anisotropy patterns by correlating shear-wave splitting and surface-wave data. *Phys. Earth Planet. Inter.* **176,** 198–212 (2009).

Author Contributions A.N. analysed the data and produced the manuscript and figures. J.W. wrote the analysis and modelling code and performed the modelling. J.W. and J-M.K. supervised the analysis and commented on the manuscript and figures. All authors discussed the results and implications at all stages.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to A.N. (andy.nowacki@bristol.ac.uk).

METHODS

S–ScS differential splitting. We measured differential shear wave splitting between S and ScS phases recorded at about 500 seismic stations in North and Central America, according to the method of ref. 10. Events of $M_w \gtrsim 5.7$ in the distance range 55°–82° were used (Supplementary Table 3), because the two phases then traverse very similar regions of the upper mantle. All data were bandpass-filtered between 0.001 Hz and 0.3 Hz to remove noise. We analysed splitting in the phases using the minimum eigenvalue technique[33], with 100 analysis windows in each case to estimate the uncertainties in φ and δt using a statistical F-test[34,35]. An example is shown in Supplementary Fig. 3. The λ_2 surfaces for measurements along each azimuth are stacked[34] in three regions (Fig. 2) to reduce the errors greatly.

Correcting for upper mantle anisotropy. We correct for upper mantle anisotropy using previously published[31,32] SKS splitting measurements (distance >90°) at stations that show little variation of splitting parameters with backazimuth, corresponding to simple upper mantle anisotropy, and where there are measurements made along similar backazimuths to the phases we measure in this study (S, ScS). These provide an estimate of the receiver-side anisotropy, and should eliminate the chance that lateral heterogeneity or dipping or multiple layers of anisotropy beneath the receiver affect our results. Analysing the splitting in S after applying a receiver-side correction gives an estimate of the source-side splitting beneath the earthquake (Fig. 1b; Supplementary Table 3). For nearby stations with no available SKS measurements, measuring splitting in S while correcting for the source anisotropy gives a receiver-side estimate. Both corrections are then applied (for shallow earthquakes; only a receiver-side correction is applied for very deep events >550 km, assuming mantle isotropy below this depth) when analysing ScS, so that the remnant splitting occurs in ScS only, and hence results from anisotropy in D″ alone. An example of a measurement where both source and receiver corrections are applied is shown in Supplementary Fig. 4.

Testing SKS splitting measurements as upper mantle anisotropy corrections. We test the validity of using SKS measurements as a correction for upper mantle anisotropy. Because the tectonic and geological processes which cause upper mantle anisotropy are unlikely to be determined by structure in D″, we can regard the two as independent. Hence over broad, continental scales, SKS measurements will be oriented approximately randomly, and we can check that the consistency observed in our results is not due to a systematic error being introduced by upper mantle anisotropy. For the Mid-Atlantic Ridge event of 2008-144-1935 (23 May), we analyse the S phase at each station for which we selected reliable SKS measurements, and replace those with others taken at random. The false 'corrections' are determined by allowing the correction fast orientation φ_{corr} to vary between 0° and 180°, and the delay time δt_{corr} between the minimum and maximum values for those in SKS measurements used in this study (0–2.5 s). A uniform random distribution is used. Supplementary Fig. 8 shows polar histograms of φ'', the projected fast orientation at the source, for five of the sets of false 'corrections'. Of these, the smallest sample standard deviation $\sigma_{\varphi''} = 47°$. Also shown is φ'' for the true SKS splitting parameters used ($\sigma_{\varphi''} = 33°$). Red bars indicate measurements of $\delta t > 3.5$ s, which may correspond to two situations. First, they may be null measurements, which frequently display a minimum at the extreme of the permitted δt (here, 4 s). These arise because by chance the 'correction' applied is the same as the total source-side and receiver splitting combined (that is, $\varphi_{SKS} \approx \varphi$ and $\delta t_{SKS} \approx \delta t$), and by removing the 'correction' there is no remnant splitting. Second, the large results may happen when the 'correction' is large and nearly perpendicular to the source and receiver splitting at the receiver, leading to a very large result that is extremely unlikely to exist in nature.

It appears that the source-side splitting direction (and also delay time; not shown) is most consistent when using SKS measurements to correct for splitting introduced after that beneath the source in S. In addition, φ'' is most similar to the plate spreading direction for the SKS-corrected case.

To confirm that applying an SKS measurement as an upper mantle splitting correction is valid, we check that particle motion is linearized and a null (or very small) measurement results from analysing an S wave from a very deep event. This confirms that the S and SKS waves undergo the same splitting while travelling in the upper mantle beneath the station, and hence that the SKS correction is valid. For the event 2007-202-1327 (21 July), Supplementary Fig. 5 shows the splitting in S at station KAPO (Kapuskasing, Ontario, Canada) with no correction applied and with the SKS measurement of ref. 36 used as a receiver correction ($\varphi_{SKS} = 69°$ and $\delta t_{SKS} = 0.58$ s). As is evident, with no correction we measure splitting in S to be the same as that in SKS within error. The removal of the splitting leads to a null result, with the particle motion highly linear (Supplementary Fig. 5d).

Source-side anisotropy estimates. A further test of the efficacy of correcting for upper mantle anisotropy with SKS measurements, after running the analyses, is to compare the source-side upper mantle splitting that remains after analysing S waves from shallow earthquakes to local splitting measurements. If there is no

contamination from unexpected or complicated anisotropy beneath the receiver for which we have not accounted, or for which SKS measurements are not an adequate correction, then source splitting parameters and local ones should be the same. For events at the East Pacific Rise, we may directly compare φ'' with measurements of SKS splitting using ocean bottom seismometers[37]. These are shown with φ'' and δt for the event 1994-246-1156 (3 September) (Supplementary Fig. 6 and Supplementary Table 3). Local splitting and splitting measured beneath the earthquake are extremely alike. This is also very strong confirmation that the source correction is a true measurement of source-side splitting, and we can thus remove it comprehensively when analysing ScS.

Source polarization measurements. Another test of the efficacy of using SKS measurements to correct for receiver-side anisotropy is to compare the predicted source polarizations of the S wave according to the Global CMT (http://www.globalcmt.org/) moment tensor solution for that event with the polarizations of the linearized particle motion after applying a correction for receiver-side upper mantle anisotropy and measuring the source-side splitting in S. For deep earthquakes, we measure the splitting in S and compare the linearized particle motion with the predicted source polarization without applying any upper mantle correction; for shallow events we apply a correction using SKS measurements. We find that in no case do the measured and predicted source polarizations differ by more than 20°, and in most cases they are within 10°. Supplementary Fig. 7 compares the predicted and measured horizontal particle motions for each earthquake used in this study at an example station.

S–ScS splitting from deep versus shallow earthquakes. As a final check that we adequately remove source-side anisotropy, we compare the results of differential analysis of S and ScS using the 2007-202-1327 (21 July) event (shown to have no measurable source anisotropy in Supplementary Fig. 5) with those from five shallow earthquakes located nearby (Supplementary Table 4). Hence, the ray paths are very similar, and the same region of D″ is sampled. If there is any systematic error in our attempt to remove the source-side splitting, the results will be significantly different.

From a larger group of 25 events located near to event 2007-202 above 100 km depth from 1989 onwards, five were selected for good signal-to-noise ratios for both S and ScS. Using $\varphi_{corr} = 70°$ and $\delta t_{corr} = 0.63$ s (average of S and SKS splitting parameters; see Supplementary Fig. 5), the procedure outlined above was conducted to obtain φ' and δt. The ray paths for measurements in the S region that traverse the region in D″ most similar to those from the shallow events were selected for comparison (Supplementary Fig. 9a). Supplementary Fig. 9c–d shows polar histograms of the fast direction in the ray frame, φ', for the two sets of results, with the near-null results downweighted in the shallow case, because the number of data points is small. Because there are few measurements, there is some spread and the standard deviation is relatively large (both of which are reduced when using larger samples; see for instance Supplementary Fig. 11, eastmost histogram). However, for the deep event, $\varphi' = 81°$ and $\langle \delta t \rangle \approx 1.3$ s; for the shallow events, $\varphi' = -84°$ and $\langle \delta t \rangle \approx 1.8$ s. Although these are not identical, they are the same within error. The small variation might be due to local variation within D″, given that the ray paths do not overlap completely. Where they do, as shown in Supplementary Fig. 9e–f, the results are the same within the 95% confidence limit, further suggesting that the difference between the two groups is mainly small local variation, not a bias in the shallow or deep source region.

This, and the other tests of the use of source and receiver corrections, compels us to believe that the shear wave splitting we observe in ScS after removing upper mantle anisotropy must be the true signal from a third, intermediate anisotropic region—D″.

Mineral slip system fitting. To compare different slip systems in ppv, we calculate the orientations of the shear planes and slip directions that are compatible with our measurements. These orientations are computed by performing a grid search over the elastic constants for the relevant slip systems[16,20,25], which are rotated about the three principal (orthogonal) axes; we scale the elastic constants by linearly mixing the fully anisotropic constants with those of an isotropic average. The amount and orientation of shear wave splitting is computed at each node using the Christoffel equation, and orientations which are compatible with the measured anisotropy (within the errors of the azimuthal stacks; Supplementary Table 4) are plotted. The larger the scaling required to fit the case, the higher degree of 'strain' is represented (indicated by colour; Fig. 3b–i), and this directly corresponds to the proportion of the material that is a linear mix of the anisotropic and isotropic components (that is, the relative proportions of oriented and random crystals).

33. Silver, P. & Chan, W. W. Shear-wave splitting and subcontinental mantle deformation. *J. Geophys. Res. Solid Earth* **96**, 16429–16454 (1991).
34. Wolfe, C. & Silver, P. Seismic anisotropy of oceanic upper mantle: Shear wave splitting methodologies and observations. *J. Geophys. Res. Solid Earth* **103**, 749–771 (1998).
35. Teanby, N., Kendall, J. M. & der Baan, M. V. Automation of shear-wave splitting measurements using cluster analysis. *Bull. Seismol. Soc. Am.* **94**, 453–463 (2004).

36. Frederiksen, A. W. *et al.* Lithospheric variations across the Superior Province, Ontario, Canada: evidence from tomography and shear wave splitting. *J. Geophys. Res. Solid Earth* **112,** B07318 (2007).

37. Wolfe, C. & Solomon, S. Shear-wave splitting and implications for mantle flow beneath the MELT region of the East Pacific Rise. *Science* **280,** 1230–1232 (1998).

LETTER

Late middle Eocene epoch of Libya yields earliest known radiation of African anthropoids

Jean-Jacques Jaeger[1], K. Christopher Beard[2], Yaowalak Chaimanee[3], Mustafa Salem[4], Mouloud Benammi[1], Osama Hlal[4], Pauline Coster[1], Awad A. Bilal[5], Philippe Duringer[6], Mathieu Schuster[1], Xavier Valentin[1], Bernard Marandat[7], Laurent Marivaux[7], Eddy Métais[8], Omar Hammuda[4] & Michel Brunet[1,9]

Reconstructing the early evolutionary history of anthropoid primates is hindered by a lack of consensus on both the timing and biogeography of anthropoid origins[1–3]. Some prefer an ancient (Cretaceous) origin for anthropoids in Africa or some other Gondwanan landmass[4], whereas others advocate a more recent (early Cenozoic) origin for anthropoids in Asia[1,2,5], with subsequent dispersal of one or more early anthropoid taxa to Africa. The oldest undoubted African anthropoid primates described so far are three species of the parapithecid *Biretia* from the late middle Eocene Bir El Ater locality of Algeria[6] and the late Eocene BQ-2 site in the Fayum region of northern Egypt[7]. Here we report the discovery of the oldest known diverse assemblage of African anthropoids from the late middle Eocene Dur At-Talah escarpment in central Libya. The primate assemblage from Dur At-Talah includes diminutive species pertaining to three higher-level anthropoid clades (Afrotarsiidae, Parapithecidae and Oligopithecidae) as well as a small species of the early strepsirhine primate *Karanisia*. The high taxonomic diversity of anthropoids at Dur At-Talah indicates either a much longer interval of anthropoid evolution in Africa than is currently documented in the fossil record or the nearly synchronous colonization of Africa by multiple anthropoid clades at some time during the middle Eocene epoch.

The chronology and biogeography of anthropoid origins have long been debated[1–7]. Molecular estimates of anthropoid origins typically advocate an early origin for the group, often extending back to the late Cretaceous[8]. In contrast, palaeontological data generally support a Cenozoic origin for anthropoids, although a wide range of potential origination dates have been suggested on the basis of fossils, of ages ranging from Palaeocene to later Eocene[1]. Similarly, there is no current consensus on where anthropoids originated. Since the discovery of a series of diverse anthropoid faunas in the Fayum region of Egypt, it has often been assumed that Africa was the birthplace of the anthropoid clade[9–11]. This interpretation has been challenged by the discovery of multiple taxa of basal anthropoids in Asia[5,12–15] and the recent finding that the putative early or middle Eocene African anthropoid *Algeripithecus* is actually a strepsirhine[16]. With the possible exception of the enigmatic *Altiatlasius koulchii* from the late Palaeocene epoch of Morocco[17], the oldest African anthropoids acknowledged so far come from the late middle Eocene (about 40 Myr ago) Bir El Ater locality in Algeria[6]. Here we augment the record of African anthropoids from the late middle Eocene on the basis of a new micromammal assemblage from Dur At-Talah in central Libya (Fig. 1). This fauna includes a small-bodied strepsirhine and a diversity of basal anthropoids, including primitive representatives of Afrotarsiidae, Parapithecidae and Oligopithecidae. The age and diversity of the Dur At-Talah primate fauna indicates substantial gaps in either the African or the Asian fossil record of anthropoid evolution (and possibly both).

The Dur At-Talah escarpment was first explored palaeontologically during the second half of the twentieth century[18]. This early phase of exploration yielded a vertebrate fauna mainly composed of taxa having medium to large body size, such as the early proboscideans *Barytherium grave*, *Arcanotherium savagei* and *Moeritherium chehbeurameuri*. Our recent fieldwork at Dur At-Talah has focused on enhancing the vertebrate record from this region by concentrating on the previously neglected microfauna. In addition to the primates reported here, five taxa of phiomyid rodents have been identified so far[19]. Biostratigraphic correlation based mainly on rodents and proboscideans suggests that the Dur At-Talah fauna approximates that from Bir El Ater in Algeria[19], which is regarded as late middle Eocene[20,21]. This correlation is supported by the new data from fossil primates described here. Available biostratigraphic evidence is also consistent with palaeomagnetic data from the Dur At-Talah section, which suggest correlation with Chron 18n.1n (38–39 Myr ago; late Bartonian)[19]. Specimens described here are housed in the palaeontological collections of Al Fateh University (Tripoli, Libya).

Primates Linnaeus, 1758
Strepsirhini Geoffroy, 1812
Lorisiformes Gregory, 1915
Karanisia Seiffert *et al.*, 2003
Karanisia arenula, sp. nov.

Holotype. DT1-42, left M_2 (Fig. 2e).
Horizon and locality. DT-Loc.1, Bioturbated Unit, Bartonian Dur At-Talah escarpment, central Libya[19].
Diagnosis. Differs from *Karanisia clarki*[22] in being smaller (adult body mass is estimated at 120–132 g). For hypodigm, description and metrics, see Supplementary Information.
Etymology. arena (Latin): sand, refers to the sandy matrix that yielded the hypodigm; -ula (Latin): diminutive suffix, in allusion to the small size of this species.

Anthropoidea Mivart, 1864
Afrotarsiidae Ginsburg and Mein, 1987
Afrotarsius Simons and Bown, 1985
Afrotarsius libycus, sp. nov.

Holotype. DT1-35, left M_1 or M_2 (Fig. 2k, l).
Horizon and locality. DT-Loc.1, Bioturbated Unit, Bartonian Dur At-Talah escarpment, central Libya[19].
Diagnosis. Differs from *Afrotarsius chatrathi*[23] in having narrower lower molars bearing hypoconid and entoconid cusps that are less isolated and less spire-like. Hypoconulid of M_1 or M_2 projects farther distally than in *A. chatrathi*. Adult body mass estimated at 130–232 g. For hypodigm, description and metrics, see Supplementary Information.
Etymology. Refers to the provenance of this species.

[1]Institut International de Paléoprimatologie et Paléontologie humaine, Évolution et Paléoenvironnements, CNRS UMR 6046, Université de Poitiers, 40 Avenue du Recteur Pineau, 86022 Poitiers, France. [2]Section of Vertebrate Paleontology, Carnegie Museum of Natural History, 4400 Forbes Avenue, Pittsburgh, Pennsylvania 15213 USA. [3]Paleontology Section, Department of Mineral Resources, Rama VI Road, Bangkok 10400, Thailand. [4]Geology Department, Al Fateh University, Tripoli, Libya. [5]Geology Department, Garyounis University, Benghazi, Libya. [6]Institut de Physique du Globe de Strasbourg, CNRS/Université de Strasbourg UMR 7516, Institut de Géologie, 1 rue Blessig, 67084 Strasbourg, France. [7]Institut des Sciences de l'Évolution, CNRS UMR 5554, Université Montpellier II, Place E. Bataillon, 34095 Montpellier, France. [8]Groupe 'TOTAL', 2 Place J. Millier, la Défense 6, 92400 Courbevoie, France. [9]Chaire de Paléontologie Humaine, Collège de France, Place M. Berthelot, 75005 Paris, France.
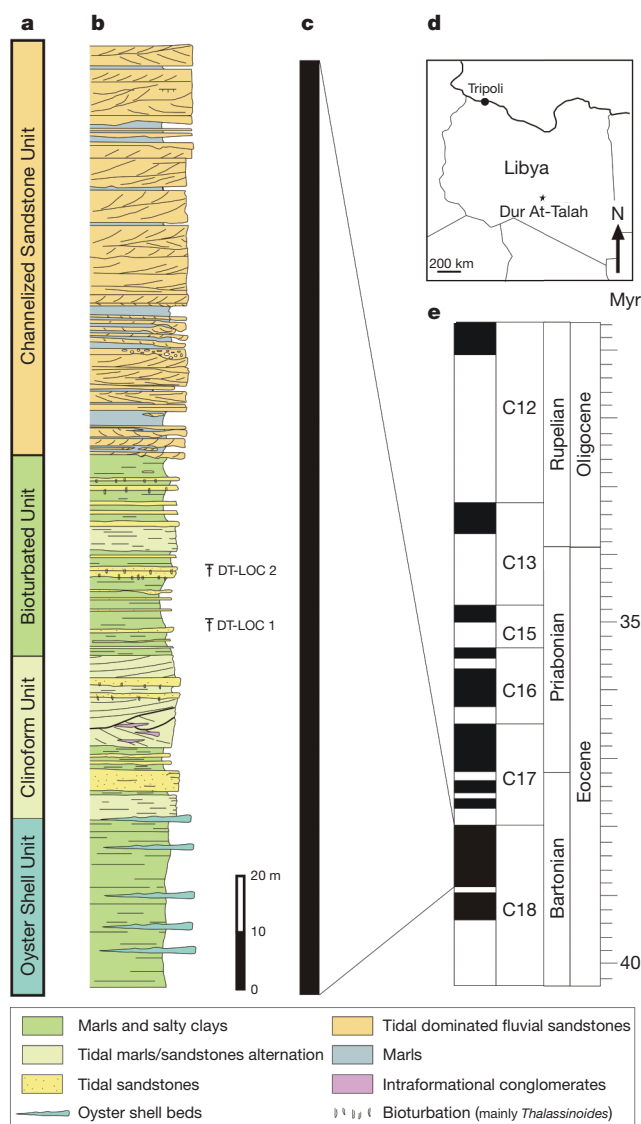
Figure 1 | Stratigraphy and correlation of the Dur At-Talah section.
a, Stratigraphic units[19]. **b**, Lithology and sedimentology of the section. **c**, Local magnetic polarity stratigraphy (black bar indicates zone of normal polarity). **d**, Map of Libya showing the geographic position of the Dur At-Talah escarpment. **e**, Preferred correlation to the Geomagnetic Polarity Time Scale[19,29].

Parapithecidae Schlosser, 1911
Biretia piveteaui de Bonis *et al.*, 1988

Referred material. DT1-26, left M[1]; DT1-27, right M[2]; DT1-28, right M[3]; DT1-29, left M[3]; DT2-23, right M[3]; DT2-24, right M[2] (Fig. 2q–w).
Horizon and locality. DT-Loc.1 and DT-Loc.2, Bioturbated Unit, Bartonian Dur At-Talah escarpment, central Libya[19].
Emended diagnosis. *Biretia piveteaui*[6] (adult body mass estimated at 292–470 g) is larger than *B. fayumensis*. M[1–2] differ from those of *B. fayumensis*[7] and *B. megalopsis*[7] in having more isolated metaconules lacking any connection with either the protocone or the metacone. M[3] mesiodistally shorter than that of *B. megalopsis*. M[3] with smaller metacone and less extensive trigon lacking metaconule, in contrast to that of *B. megalopsis*. For description and metrics, see Supplementary Information.

Oligopithecidae Simons, 1989
Talahpithecus parvus, gen. et sp. nov.

Holotype. DT1-31, left M[1] or M[2] (Fig. 2n).

Horizon and locality. DT-Loc.1, Bioturbated Unit, Bartonian Dur At-Talah escarpment, central Libya[19].
**Diagnosis.** Smaller (adult body mass estimated at 226–376 g) than *Catopithecus* and *Oligopithecus*. Upper molars without mesostyle and with smaller hypocone than in *Catopithecus*. Crests surrounding upper molar trigon more trenchant than in *Oligopithecus* and *Catopithecus*. Lower molars with relatively narrower talonid and higher trigonid with more nearly vertical postvallid than in *Oligopithecus* and *Catopithecus*. For hypodigm, description and metrics, see Supplementary Information.
Etymology. talah (Arabic): tree, refers to the provenance of this genus; parvus (Latin): small, refers to the size of this species.

All four primate taxa currently known from Dur At-Talah are remarkably small, ranging from 120 to 470 g in estimated adult body mass. Such a small size distribution for the earliest known African radiation of anthropoids reinforces the conclusion drawn from analysis of the middle Eocene primate assemblage of Shanghuang, China, that the origin of anthropoids occurred at very small body size[24]. Indeed, if recent phylogenetic analyses recognizing oligopithecids as early members of the catarrhine clade are correct[7], the small size of *Talahpithecus parvus* would suggest that even the origin of crown anthropoids and the platyrrhine/catarrhine divergence occurred at small body mass. However, by the time of the late Eocene L-41 primate fauna from the Fayum region of Egypt[10], larger anthropoid taxa had begun to supplant these diminutive taxa, and this trend towards increasing body mass among early African anthropoids continued into the Oligocene epoch. The common occurrence of *Biretia piveteaui* at both Bir El Ater and Dur At-Talah supports a similar age for these faunas. The small size of *Karanisia arenula* from Dur At-Talah in comparison with *K. clarki* from BQ-2 in the Fayum, as well as the small size and primitive anatomy of *Talahpithecus parvus* in comparison with Fayum oligopithecids such as *Catopithecus browni*, reinforce biostratigraphic data from rodents and proboscideans suggesting that Dur At-Talah is roughly equivalent to Bir El Ater in age. Both of the latter faunas seem to be older than BQ-2 in the Fayum[19].

The phylogenetic affinities of three of the four primate taxa documented at Dur At-Talah are uncontroversial, but there is no current consensus regarding the broader affinities of *Afrotarsius*, represented at Dur At-Talah by *A. libycus*. Originally described as a possible African tarsiid (hence the generic name)[23], multiple subsequent authors have suggested that *Afrotarsius* is a basal member of the anthropoid clade[9,25,26]. The previously unknown upper-molar morphology of *Afrotarsius*, documented here, supports an attribution of this genus to Anthropoidea rather than Tarsiidae (or Tarsiiformes). Like those of Asian eosimiid anthropoids (*Eosimias*, *Phenacopithecus* and *Bahinia*)[5,14], the upper molars of *Afrotarsius* bear an elongated postmetacrista and an enlarged shelf-like structure buccal to the metacone. The upper molars of *Afrotarsius* and eosimiids also share transversely oriented crests that variably connect the paracone and metacone with their associated conules (or remnants thereof). The upper molars of *Afrotarsius* differ from those of eosimiids in retaining continuity between the postmetaconule crista and the postcingulum, which is lost in eosimiids. As noted by previous authors[25], M[3] of *Afrotarsius* is distinctively anthropoid-like (and differs from that of tarsiids) in having a remarkably abbreviated hypoconulid lobe (Fig. 2m). In view of these anatomical characters, we regard *Afrotarsius* as a relatively basal member of the anthropoid clade. However, substantial additional evidence will be required to ascertain how *Afrotarsius* relates to other early anthropoid taxa, particularly eosimiids. Dental similarities between *Afrotarsius* and tarsiids probably reflect the convergent acquisition of trenchant molar crests as an adaptation for insectivory.

The presence of three distinct clades of anthropoids (Afrotarsiidae, Parapithecidae and Oligopithecidae) in the late middle Eocene Dur At-Talah fauna is surprising, especially in view of the lower diversity of early anthropoids that has been described so far from the BQ-2 locality of late Eocene age in northern Egypt[7]. Recent comprehensive analyses of early anthropoid relationships disagree on many aspects of tree
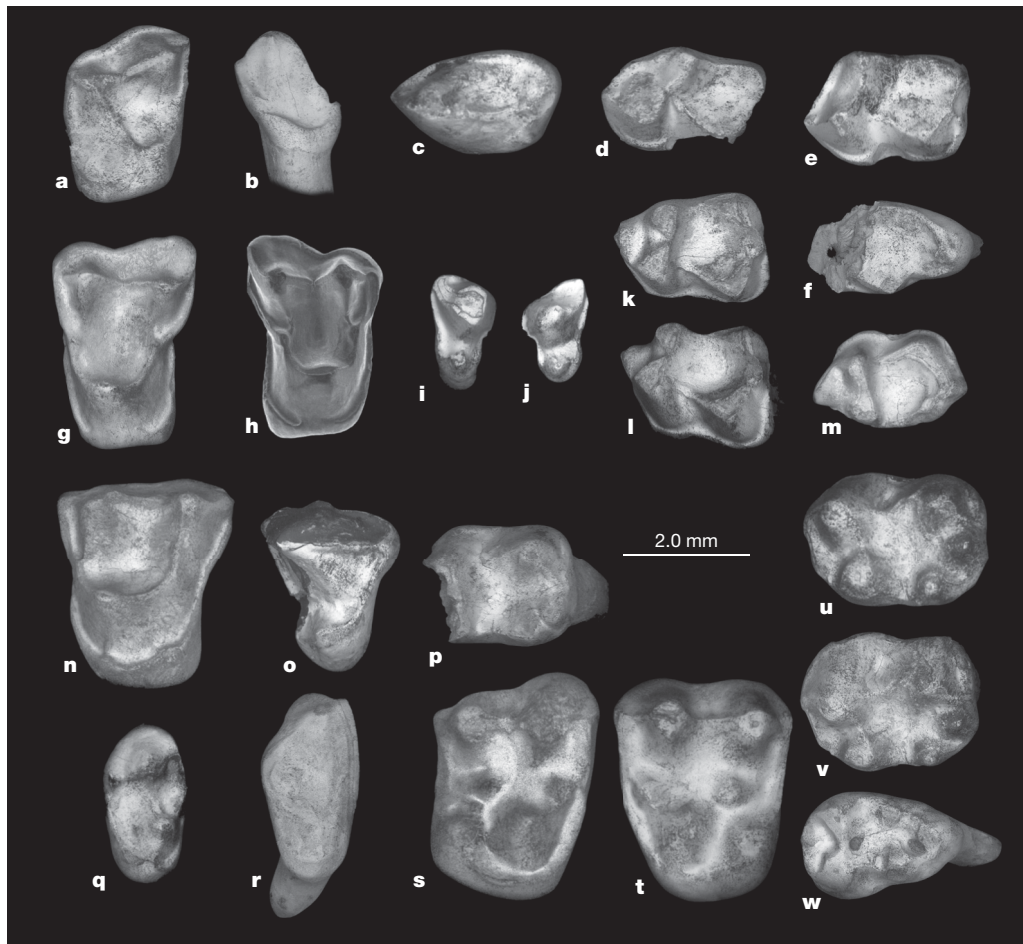
Figure 2 | Scanning electron microscope images of fossil primate teeth from Dur At-Talah. a–f, *Karanisia arenula* sp. nov. **a,** Right M^3 (DT1-37), occlusal view. **b,** Right P_3 (DT1-38), lingual view. **c,** Left P_4 (DT1-39), occlusal view. **d,** Left M_1 (DT1-41), occlusal view. **e,** Holotype left M_2 (DT1-42), occlusal view. **f,** Fragmentary left M_3 (DT1-43), occlusal view. **g–m,** *Afrotarsius libycus* sp. nov. **g,** Left M^2 (DT1-33), occlusal view. **h,** Right M^2 (DT1-34), occlusal view. **i,** Right P^3 (DT1-31), occlusal view. **j,** Left P^3 (DT1-32), occlusal view. **k,** Holotype left M_1 or M_2 (DT1-35), occlusal view. **l,** Holotype left M_1 or M_2 (DT1-35), oblique buccal view. **m,** Right M_3 (DT1-36), occlusal view. **n–p,** *Talahpithecus parvus* gen. et sp. nov. **n,** Holotype left M^1 or M^2 (DT1-31), occlusal view. **o,** Right P^4 (DT1-30), mesial oblique view. **p,** Fragmentary right M_1 or M_2 (DT1-32), occlusal view. **q–w,** *Biretia piveteaui*. **q,** Right M^3 (DT2-23), occlusal view. **r,** Right M^3 (DT1-28), occlusal view. **s,** Right M^2 (DT1-27), occlusal view. **t,** Left M^1 (DT1-26), occlusal view. **u,** Right M_2 (DT2-24), occlusal view. **v,** Right M_2 (DT2-24), oblique buccal view. **w,** Left M_3 (DT1-29), occlusal view.

topology[7,27], but all current reconstructions of early anthropoid phylogeny insist that the three anthropoid clades represented at Dur At-Talah occupy disparate positions on the evolutionary tree. The high degree of morphological, taxonomic and presumably ecological diversity apparent in the Dur At-Talah anthropoid fauna can be explained only by a substantial interval of earlier evolutionary history for this group. Given the apparent absence of anthropoids in significantly older, but reasonably well sampled, Eocene African localities such as Glib Zegdou in western Algeria[16], it seems doubtful that the 'missing' evolutionary history of the Dur At-Talah anthropoids can be explained simply by reference to the poorly sampled early Cenozoic fossil record of Africa. An alternative hypothesis that now demands serious consideration is that multiple Asian anthropoid clades may have colonized Africa more or less synchronously during the middle Eocene, alongside anomaluroid and hystricognathous rodents. In either case, further palaeontological exploration of middle Eocene localities in Africa and Asia will be necessary to illuminate this poorly documented interval of primate evolutionary history.

METHODS SUMMARY

Taxonomic allocation. Fossil specimens from Dur At-Talah were segregated into taxa on the basis of both metric and morphological compatibility. Specimens from Dur At-Talah were extensively compared with original specimens and casts of African and Asian fossil primates to establish the systematic affinities of the Dur At-Talah taxa.

Estimation of body mass. Mean estimates of adult body mass for each primate taxon from Dur At-Talah were obtained by using the regression equations provided by Conroy[28]. Conroy's regressions estimate body mass on the basis of M_1 area. This tooth locus is not definitively known for any of the Dur At-Talah anthropoid taxa, because M_1 and M_2 are not readily distinguished in *Afrotarsius* and because the sole lower molar currently known for *Talahpithecus parvus* is fragmentary (see Supplementary Information). In these cases, M_2 dimensions may have been substituted for M_1 (as was certainly the case for *Biretia piveteaui*). Two regression equations were used to estimate adult body mass for each primate taxon known from Dur At-Talah. Conroy's 'all primates' regression was used in every case, although more taxonomically restricted regressions were also employed (Conroy's 'prosimians' regression was used for *Karanisia*, and Conroy's 'monkeys' regression was used for the anthropoids).

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 7 July; accepted 16 August 2010.

1. Beard, K. C. *The Hunt for the Dawn Monkey: Unearthing the Origins of Monkeys, Apes, and Humans* (Univ. California Press, Berkeley, 2004).
2. Beard, K. C. in *Primate Biogeography* (eds Lehman, S. G. & Fleagle, J. G.) 439–467 (Springer, 2006).
3. Williams, B. A., Kay, R. F. & Kirk, E. C. New perspectives on anthropoid origins. *Proc. Natl Acad. Sci. USA* **107,** 4797–4804 (2010).
4. Miller, E. R., Gunnell, G. F. & Martin, R. D. Deep time and the search for anthropoid origins. *Yearb. Phys. Anthropol.* **48,** 60–95 (2005).
5. Jaeger, J.-J. *et al.* A new primate from the middle Eocene of Myanmar and the Asian early origin of anthropoids. *Science* **286,** 528–530 (1999).

6. de Bonis, L., Jaeger, J.-J., Coiffait, B. & Coiffait, P.-E. Découverte du plus ancien primate catarrhinien connu dans l'Éocène supérieur d'Afrique du Nord. *C. R. Acad. Sci.* **306,** 929–934 (1988).

7. Seiffert, E. R. *et al.* Basal anthropoids from Egypt and the antiquity of Africa's higher primate radiation. *Science* **310,** 300–304 (2005).

8. Bininda-Emonds, O. R. P. *et al.* The delayed rise of present-day mammals. *Nature* **446,** 507–512 (2007).

9. Fleagle, J. G. & Kay, R. F. The phyletic position of the Parapithecidae. *J. Hum. Evol.* **16,** 483–532 (1987).

10. Simons, E. L. Diversity in the early Tertiary anthropoidean radiation in Africa. *Proc. Natl Acad. Sci. USA* **89,** 10743–10747 (1992).

11. Ciochon, R. L. & Gunnell, G. F. Chronology of primate discoveries in Myanmar: influences on the anthropoid origins debate. *Yearb. Phys. Anthropol.* **45,** 2–35 (2002).

12. Beard, K. C., Qi, T., Dawson, M. R., Wang, B.-Y. & Li, C.-K. A diverse new primate fauna from middle Eocene fissure-fillings in southeastern China. *Nature* **368,** 604–609 (1994).

13. Beard, K. C., Tong, Y.-S., Dawson, M. R., Wang, J.-W. & Huang, X.-S. Earliest complete dentition of an anthropoid primate from the late middle Eocene of Shanxi Province, China. *Science* **272,** 82–85 (1996).

14. Beard, K. C. & Wang, J.-W. The eosimiid primates (Anthropoidea) of the Heti Formation, Yuanqu Basin, Shanxi and Henan Provinces, People's Republic of China. *J. Hum. Evol.* **46,** 401–432 (2004).

15. Beard, K. C. *et al.* A new primate from the Eocene Pondaung Formation of Myanmar and the monophyly of Burmese amphipithecids. *Proc. R. Soc. Lond. B* **276,** 3285–3294 (2009).

16. Tabuce, R. *et al.* Anthropoid versus strepsirhine status of the African Eocene primates *Algeripithecus* and *Azibius*: craniodental evidence. *Proc. R. Soc. Lond. B* **276,** 4087–4094 (2009).

17. Sigé, B., Jaeger, J.-J., Sudre, J. & Vianey-Liaud, M. *Altiatlasius koulchii* n. gen. et sp., primate omomyidé du Paléocène supérieur du Maroc, et les origines des euprimates. *Palaeontographica A* **214,** 31–56 (1990).

18. Wight, A. W. R. in *The Geology of Libya* Vol. 1 (eds Salem, M. J. & Busrewil, M. T.) 309–325 (Academic, 1980).

19. Jaeger, J.-J. *et al.* New rodent assemblages from the Eocene Dur At-Talah escarpment (Sahara of central Libya): systematic, biochronological, and palaeobiogeographical implications. *Zool. J. Linn. Soc.* **160,** 195–213 (2010).

20. Tabuce, R., Coiffait, B., Coiffait, P.-E., Mahboubi, M. & Jaeger, J.-J. A new species of *Bunohyrax* (Hyracoidea, Mammalia) from the Eocene of Bir el Ater (Algeria). *C. R. Acad. Sci.* **331,** 61–66 (2000).

21. Tabuce, R., Coiffait, B., Coiffait, P.-E., Mahboubi, M. & Jaeger, J.-J. A new genus of Macroscelidea (Mammalia) from the Eocene of Algeria: a possible origin for elephant-shrews. *J. Vertebr. Paleontol.* **21,** 535–546 (2001).

22. Seiffert, E. R., Simons, E. L. & Attia, Y. Fossil evidence for an ancient divergence of lorises and galagos. *Nature* **422,** 421–424 (2003).

23. Simons, E. L. & Bown, T. M. *Afrotarsius chatrathi*, first tarsiiform primate (?Tarsiidae) from Africa. *Nature* **313,** 475–477 (1985).

24. Gebo, D. L., Dagosto, M., Beard, K. C. & Qi, T. The smallest primates. *J. Hum. Evol.* **38,** 585–594 (2000).

25. Ginsburg, L. & Mein, P. *Tarsius thailandica* nov. sp., premier Tarsiidae (Primates, Mammalia) fossile d'Asie. *C. R. Acad. Sci.* **304,** 1213–1215 (1987).

26. Beard, K. C. A new genus of Tarsiidae (Mammalia: Primates) from the middle Eocene of Shanxi Province, China, with notes on the historical biogeography of tarsiers. *Bull. Carnegie Mus. Nat. Hist.* **34,** 260–277 (1998).

27. Kay, R. F., Williams, B. A., Ross, C. F., Takai, M. & Shigehara, N. in *Anthropoid Origins: New Visions* (eds Ross, C. F. & Kay, R. F.) 91–135 (Kluwer/Plenum, 2004).

28. Conroy, G. C. Problems of body-weight estimation in fossil primates. *Int. J. Primatol.* **8,** 115–137 (1987).

29. Gradstein, F. M., Ogg, J. G. & Smith, A. G. *A Geological Time Scale 2004* (Cambridge Univ. Press, 2004).

Author Contributions M.Br., M.Sa., O.H. and J.-J.J. designed and organized the project. J.-J.J., Y.C., X.V., M.Be., M.S., A.A.B., B.M. and L.M. collected palaeontological data. M.Be., P.C., M.Sch, O.Hl, P.D. and E.M. collected geological and palaeomagnetic data. J.-J.J., Y.C. and K.C.B. analysed the data. J.-J.J. and K.C.B. wrote the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to J.-J.J. (jean-jacques.jaeger@univ-poitiers.fr)

METHODS

Taxonomic allocation. Fossil specimens from Dur At-Talah were segregated into taxa on the basis of both metric and morphological compatibility. The following taxa of Eocene–Oligocene primates from Africa and Asia formed the comparative sample used to make taxonomic decisions regarding the Dur At-Talah primates: *Karanisia clarki, Saharagalago misrensis, Tarsius eocaenus, Xanthorhysis tabrumi, Afrotarsius chatrathi, Eosimias sinensis, E. centennicus, E. dawsonae, Phenacopithecus krishtalkai, P. xueshii, Bahinia pondaungensis, Biretia piveteaui, B. fayumensis, B. megalopsis, Qatrania wingi, Arsinoea kallimos, Serapia eocaena, Proteopithecus sylviae, Catopithecus browni, Oligopithecus rogeri.*

Measurements. Standard measurements (mesiodistal length, buccolingual width; separate width measurements for lower molar trigonids and talonids) were obtained for each tooth in the current sample (Supplementary Table 1). Measurements were taken to the nearest 0.01 mm with digital calipers. Equivalent dimensions were estimated in the case of two fragmentary specimens (DT1-32 and DT1-43).

Body mass estimation. Estimates of adult body mass for each primate taxon from Dur At-Talah were obtained by using the regression equations provided by Conroy[28]. Conroy's regressions estimate body mass on the basis of M_1 area. For *Karanisia arenula* body mass was estimated from the mean M_1 area of the two available specimens (DT1-40 and DT1-41). Two estimates of the adult body mass of *Karanisia arenula* were obtained, using Conroy's 'all primates' and 'prosimians' regressions, respectively. The body mass of *Afrotarsius libycus* was estimated from the dimensions of the holotype lower molar (DT1-35), which is either an M_1 or an M_2. The body mass of *Biretia piveteaui* was estimated on the basis of DT2-24, regarded here as an M_2. These teeth do not differ appreciably in size in *Afrotarsius chatrathi*[23] and Fayum species of *Biretia*[7], suggesting that any error introduced by substituting the dimensions of M_2 for those of M_1 here is negligible. Body mass of *Talahpithecus parvus* was assessed on the basis of DT1-32, a fragmentary M_1 or M_2 whose length can only be estimated because of breakage. Two estimates of the adult body mass of each of the three anthropoid taxa represented at Dur At-Talah were obtained, using Conroy's 'all primates' and 'monkeys' regressions, respectively.

LETTER

# Fine-scale recombination rate differences between sexes, populations and individuals

Augustine Kong[1], Gudmar Thorleifsson[1], Daniel F. Gudbjartsson[1], Gisli Masson[1], Asgeir Sigurdsson[1], Aslaug Jonasdottir[1], G. Bragi Walters[1], Adalbjorg Jonasdottir[1], Arnaldur Gylfason[1], Kari Th. Kristinsson[1], Sigurjon A. Gudjonsson[1], Michael L. Frigge[1], Agnar Helgason[1,2], Unnur Thorsteinsdottir[1,3] & Kari Stefansson[1,3]

Meiotic recombinations contribute to genetic diversity by yielding new combinations of alleles. Recently, high-resolution recombination maps were inferred from high-density single-nucleotide polymorphism (SNP) data using linkage disequilibrium (LD) patterns that capture historical recombination events[1,2]. The use of these maps has been demonstrated by the identification of recombination hotspots[2] and associated motifs[3], and the discovery that the *PRDM9* gene affects the proportion of recombinations occurring at hotspots[4–6]. However, these maps provide no information about individual or sex differences. Moreover, locus-specific demographic factors like natural selection[7] can bias LD-based estimates of recombination rate. Existing genetic maps based on family data avoid these shortcomings[8], but their resolution is limited by relatively few meioses and a low density of markers. Here we used genome-wide SNP data from 15,257 parent–offspring pairs to construct the first recombination maps based on directly observed recombinations with a resolution that is effective down to 10 kilobases (kb). Comparing male and female maps reveals that about 15% of hotspots in one sex are specific to that sex. Although male recombinations result in more shuffling of exons within genes, female recombinations generate more new combinations of nearby genes. We discover novel associations between recombination characteristics of individuals and variants in the *PRDM9* gene and we identify new recombination hotspots. Comparisons of our maps with two LD-based maps inferred from data of HapMap populations of Utah residents with ancestry from northern and western Europe (CEU) and Yoruba in Ibadan, Nigeria (YRI) reveal population differences previously masked by noise and map differences at regions previously described as targets of natural selection.

To perform a large, family-based recombination study, one challenge is to phase the genotypes of the parents when the grandparents are not genotyped. One solution is to use genotyped nuclear families with two or more offspring, which in essence uses the children to phase the parents. However, resolution can be diminished and difficulties can arise when two or more offspring have recombinations that are close to each other. We capitalized on recent methodological advances that led to the successful determination of parental origins of over 97% of the heterozygous genotypes of 38,167 Icelanders typed on Illumina SNP arrays, many of them with ungenotyped parents[9,10]. Parental origins provide phase. We used phased haplotypes of 8,850 mother–offspring pairs (6,041 distinct mothers) and 6,407 father–offspring pairs (4,389 distinct fathers) to identify recombinations (Fig. 1) for 15,257 meioses (Supplementary Table 1).

Recombinations were determined using 289,658 and 8,411 SNPs on the autosomal and X chromosomes respectively. The data only allowed us to assign a recombination to the region spanned by the two closest flanking heterozygous markers in the parent (Fig. 1). Treating this as a missing-data problem, the EM algorithm[11] was used to calculate likelihood-based estimates of recombination rates for males and females (Supplementary Information and Supplementary Table 2). Also, results from the E-step of the EM algorithm were used to calculate the estimated recombination count in each marker interval for each meiosis. In addition to genetic distances between SNPs, maps for various uniformly spaced grids were calculated by linear interpolation.

Existing genetic maps include the 2002 deCode family-based map[8] and the most commonly used LD-based maps[12] (Methods Summary), referred to here as the CEU, YRI and COMBINED maps. The COMBINED map is essentially the average of the CEU and YRI maps.
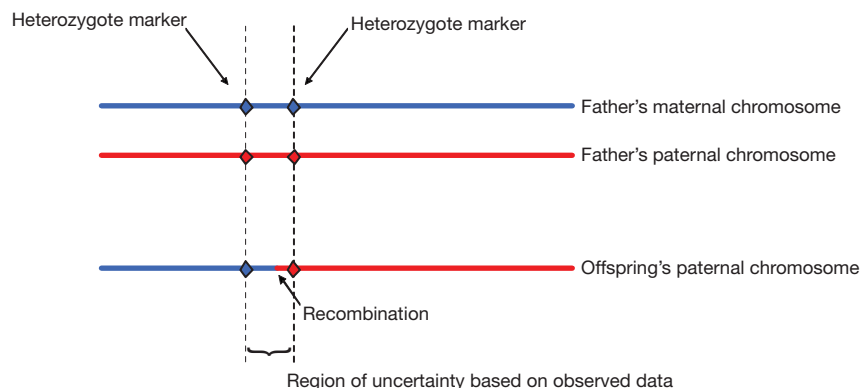
Figure 1 | Determining recombination locations. Here it is assumed that genotypes of parent and offspring have been phased, with parental origin determined[9,10]. The parent shown is a father, but the same method applies to a mother–offspring pair. For an SNP that is heterozygous for the parent, it can be determined whether the allele passed on to the offspring is from the parent's maternal or paternal chromosome. The location of the recombination can hence be localized to the region spanned by the two closest flanking heterozygous markers in the parent. (Details are in Supplementary Information.)

[1]deCODE genetics, Sturlugata 8, 101 Reykjavík, Iceland. [2]Department of Anthropology, University of Iceland, Sæmundargötu 2, 101 Reykjavik, Iceland. [3]Faculty of Medicine, University of Iceland, Sæmundargötu 2, 101 Reykjavik, Iceland.

These maps have similar lengths, because the 2002 deCode map was used to scale the other maps which only provide information about relative recombination rates. By comparison, our newly constructed sex-averaged map is 3% shorter. This is probably because we tabulated only recombinations considered highly reliable and some recombinations were missed. Also, the 2002 deCode map could be slightly inflated because of genotyping errors. Supporting the assumption that the dropped recombinations are approximately randomly distributed and have minimal impact on the relative recombination rate estimates, the correlation between the sex-averaged map and the 2002 deCode Map is 0.945 at 3-megabase (Mb) resolution, roughly the limit of resolution for the older map. This correlation is stronger than that between the 2002 deCode map and the LD-based maps ($r = 0.920$, 0.914 and 0.927 for the CEU, YRI and COMBINED maps, respectively). Correlation between the sex-averaged map and the COMBINED map is 0.977 (Supplementary Table 3).

Recombination maps partitioned into 10-kb bins were calculated for each sex. For subsequent investigations, we excluded the X chromosome and 5-Mb regions at the ends of autosomal chromosomes relative to the SNP coverage, locations where the determination of recombinations is less reliable. We also excluded 10,254 bins covering unsequenced regions (Human Map build 36), of which 8,891 were centromeric. These bins generally have low recombination rates and a fraction of them include intervals without recombination rates assigned by the COMBINED-map. Genetic distances of those bins are clearly biased downwards in all three LD-based maps. In total, the studied regions covered 2,444.46 Mb or 244,446 bins. For these bins, the estimated average genetic distance is 0.0155 cM (sum = 3,790.1 cM) for females and 0.0077 cM (sum = 1886.7 cM) for males. At this resolution, the correlation between the male and female maps is 0.659.

A standardized recombination rate (SRR) was calculated for males and females separately, by dividing the genetic distance of each bin by the overall average. Defining recombination hotspots as those bins with an SRR greater than 10, we observed 4,762 hotspots for males and 4,129 hotspots for females, with an overlap of 1,953. The male hotspot bins covered 1.9% of the physical distance of the studied region but accounted for 36.2% of its recombinations. Corresponding numbers for females were 1.7% and 28.0%, respectively. Despite similarities of sexes, 718 and 125 of the 4,762 male hotspots have an SRR less than 3 and 1, respectively, in females. A permutation test (Supplementary Information) showed that these male-specific hotspots have a false-discovery rate of approximately 1.9% and 0%, respectively. Thus approximately 704 of the 718, and all of the 125, identified bins correspond to true sex differences, indicating that about 14.8% (704/4,762) of the male hotspots are sex specific. Correspondingly, of the 4,129 female hotspots, 624 (false-discovery rate 2.8%) and 166 (false-discovery rate 0.7%) have an SRR smaller than 3 and 1, respectively, in males. About 14.7% (606/4,129) of female hotspots are sex specific.

Sex-specific hotspots tend to occur in clusters. Fig. 2a shows a region harbouring the *Basonuclin-2* gene[13] where recombinations are dominated by those resulting from male meioses. This region contains five male-specific hotspots, the two most striking being at 16.649 Mb (male SRR = 29.1) and 16.829 Mb (male SRR = 27.7). However, even though the female SRR is substantially smaller for these two bins (0.5 and 2.6, respectively), they do correspond to local peaks. This is typical for other male-specific hotspots. The same trend applies to regions where recombinations are dominated by females: that is, local peaks for male recombination rate at female-specific hotspots (Supplementary Fig. 1). Thus, even though hotspots are defined for narrow intervals (noting that the 10-kb resolution hotspots examined here could often be driven by intervals much shorter in length), they are determined by interactions between factors both local and regional, the latter concerning regions that are hundreds of kilobases to many megabases in length (Fig. 2b). If the local factors, but not the regional ones, are supportive of recombination, a local peak that is not a hotspot would result. Moreover, the regional forces influencing male and females are only
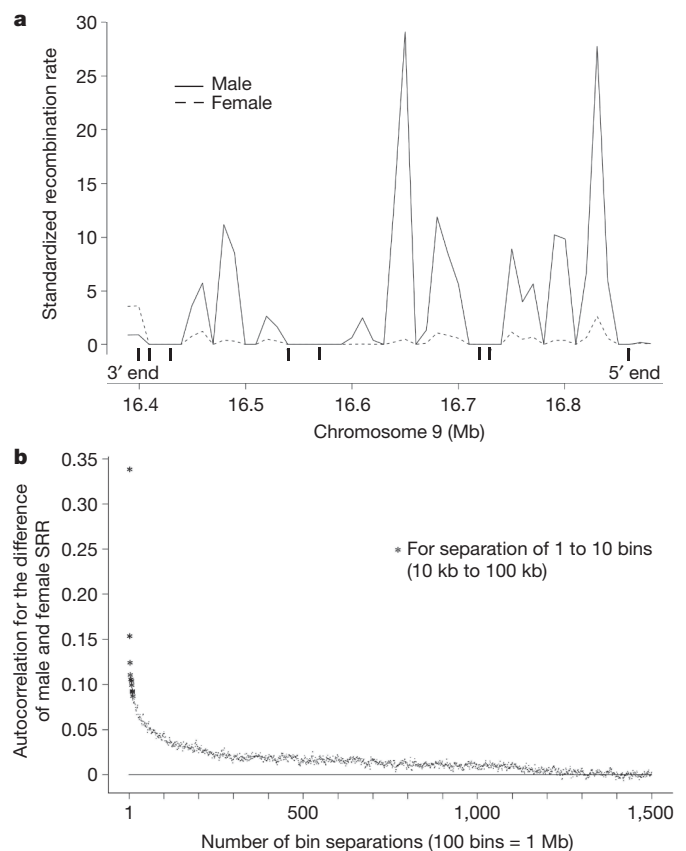
Figure 2 | Sex differences in recombinations. a, Male and female SRRs at the *Basonuclin-2* gene. Recombinations in this region are dominated by those resulting from male meioses. It is, however, noted that, although female recombination rates are generally low here, locations of male recombination hotspots often correspond to local peaks for female recombination rate. **b**, Autocorrelations of the difference between male and female SRR as a function of the number of bin separations. Note that, albeit small ($\sim$0.007), the correlation is clearly positive for bins that are 10 Mb (1,000 bins) apart.

partly correlated. Indeed, the correlation between the male and female maps is less at 3-Mb (0.649) than at 10-kb resolution (0.659), even though the former is less affected by sampling variation.

We classified the 10-kb bins as genic, intergenic or at gene boundaries (Table 1). On average, the recombination rate is lower in genic regions than in intergenic ones, a difference that is greater for females (average SRR = 0.898 and 1.053, respectively) than males (average SRR = 0.992 and 1.012, respectively). For both sexes, the recombination rate tends to be lower at genic bins containing exons, and higher for those containing only introns, particularly those where the closest exon is more than three bins away. This latter difference is much greater for males (SRR = 0.868 and 1.284, respectively) than females (SRR = 0.843 and 1.013, respectively). In fact, intron bins far from exons exhibit the greatest difference between male and female SRR (0.270, $P = 2.2 \times 10^{-7}$) among the bin categories studied. At intergenic regions, for both sexes, the recombination rate first increases with distance from the first or last exon of genes, peaking at approximately three to four bins away, then decreases. The changes are more dramatic in females than males (Table 1). For intergenic bins that are ten bins or less from genes, the average SRR for males and females is 1.119 and 1.256, respectively ($P = 1.3 \times 10^{-19}$). Hence, although more male recombinations participate in shuffling exons within genes, female meioses are characterized more by gene shuffling.

For both sexes, similar differences in SRR exist between the 5′ and 3′ ends of genes. For intergenic regions within 100 kb of the nearest gene, the average SRR at the 5′ ends is approximately 0.15 lower than that at the 3′ ends ($P = 2.7 \times 10^{-7}$). The difference disappears for distances greater than 100 kb. In contrast, bins containing the first exon of a gene

Table 1 | Sex-specific standardized recombination rate and genomic regions

Bin type (10 kb)	Number of bins	Male recombination rate§	Female recombination rate§	Male recombination rate − female recombination rate (P value)‖
Exon* (>99% genic)	40,538	0.868	0.843	0.024 (0.25)
Intron (n.e.† = 1)	20,535	1.010	0.903	0.106 (1.5×10^{-6})
Intron (n.e. = 2)	8,718	1.047	0.958	0.089 (0.0029)
Intron (n.e. = 3)	5,009	1.095	0.919	0.177 (3.9×10^{-5})
Intron (n.e. ≥ 4)	12,580	1.284	1.013	0.270 (2.2×10^{-7})
Intron (all)	46,842	1.099	0.945	0.155 (1.1×10^{-11})
Genic (exon + intron)	87,380	0.992	0.898	0.094 (1.9×10^{-8})
Gene boundary‡	24,498	0.963	1.079	−0.117 (1.3×10^{-5})
Intergenic (n.e. = 1)	15,912	1.120	1.196	−0.075 (0.0025)
Intergenic (n.e. = 2)	10,935	1.118	1.242	−0.124 (2.5×10^{-6})
Intergenic (n.e. = 3)	8,245	1.149	1.356	−0.207 (4.8×10^{-12})
Intergenic (n.e. = 4)	6,495	1.175	1.349	−0.174 (6.6×10^{-7})
Intergenic (n.e. = 5)	5,385	1.138	1.291	−0.153 (4.6×10^{-5})
Intergenic (n.e. = 6)	4,602	1.142	1.277	−0.135 (0.00062)
Intergenic (n.e. = 7)	4,017	1.126	1.242	−0.116 (0.010)
Intergenic (n.e. ≤ 10)	65,380	1.119	1.256	−0.136 (1.3×10^{-19})
Intergenic (n.e. ≥ 20)	48,057	0.845	0.773	0.072 (0.011)
Intergenic (all)	132,568	1.012	1.053	−0.041 (0.00092)

* An exon bin is one that is more than 99% genic and contains either part or full exons. †n.e., number of bins to nearest exon. ‡A gene boundary bin is one that is partly genic (that is, it contains either part of or the full first or last exon of genes) and partly intergenic (≥1%). §Recombination rates are in standardized units; that is, over the 244,446 10-kb bins studied, they average to 1 for males and females, respectively. ‖P values have been adjusted for correlation among bins close to each other using a randomization procedure (see Supplementary Information).

have a higher average SRR than those containing the last exon (~0.11, $P = 7.2 \times 10^{-4}$). This difference, however, does not extend further into the genic regions, which suggests that the first intron has a higher recombination rate than the last intron at the immediate neighbourhoods of the first and last exons. Figure 3 summarizes the relationships between sex-specific recombination rate and genes.

Differences in recombination rate exist between individuals of the same sex[8,14,15]. Recently, the *PRDM9* gene was shown to be a major determinant of hotspots in humans[4–6]. This gene is highly polymorphic, with most of its sequence variants clustering in the zinc-finger domain of the gene. The Human Genome assembly (hg18) ascribes 13 zinc-finger repeats to the *PRDM9* gene. These repeats are invariant except at positions −1, 3 and 6 of each of the zinc-finger α-helixes. Variations in the number of repeats within the human population have been described. In the Hutterite population, carriers of a rare version of the gene with 16 zinc-finger repeats were shown to have substantially fewer recombinations in hotspots than non-carriers[4]. To investigate comprehensively variants that could affect hotspots, we performed a genome scan, separately, for the 6,041 mothers and 4,389 fathers studied, correlating the fraction of recombinations in hotspots (henceforth referred to as the hotspot phenotype) with SNPs on the Illumina 1M chip (Methods Summary).
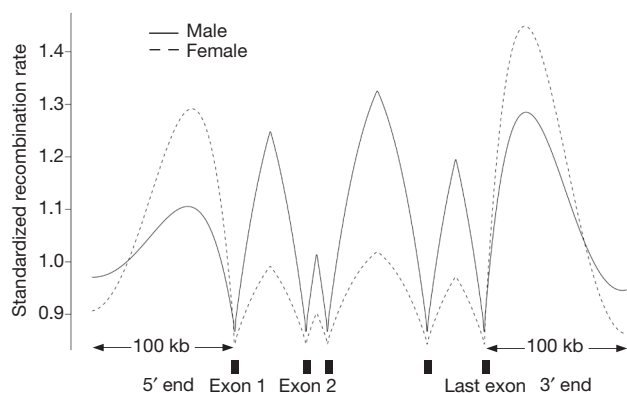
Figure 3 | Sex-specific recombination rates and genes. Schematic picture summarizing general trends (see Table 1); it is not meant to reflect the recombination rate pattern around a specific gene. Male SRR, although low at exons, tends to be high at intronic regions that are distant from exons. Male and female SRRs both tend to be high at intergenic regions around 40 kb from the first or last exon of a gene, but it is higher for females. Also, for both sexes, intergenic regions close to 3′ ends tend to have higher recombination rates than those close to 5′ ends.

For both sexes, many SNPs around *PRDM9* achieved genome-wide significance ($P < 5 \times 10^{-8}$, Supplementary Figs 2 and 3). Most significant was rs2914276, with the minor allele G (frequency = 3.9%) associating with fewer recombinations in hotspots ($P < 10^{-100}$ for females and $P < 10^{-50}$ for males). Determining the zinc-finger repeat number of 575 Icelanders, enriching for carriers of rs2914276-G, we observed variations in repeat numbers from 12 to 15 (Supplementary Information, Supplementary Fig. 4 and Supplementary Table 4). Imputing this polymorphism into others, 12 to 15 repeats were estimated to have frequencies of 0.1%, 96.0%, 3.2% and 0.6%, respectively. Rs2914276-G correlates substantially with either 14 or 15 repeats ($r = 0.83$), whereas the major allele A is correlated with 12 or 13 repeats. No significant difference was seen between12 and 13 repeats with respect to hotspots. Individuals carrying only 12 or 13 repeats have 28.6% and 37.1% of their recombinations in hotspots for females and males, respectively. Fourteen and 15 repeats are associated with significantly lower fractions. One copy of 14 repeats brings the corresponding fractions down to 19.1% and 25.4%, and for 15 repeats to 20.5% and 27.3%. Although the higher fractions for the 15 repeats than the 14 repeats are barely significant when results from both sexes are combined ($P = 0.018$), it emphasizes that the fraction of recombinations in hotspots does not decrease monotonically with number of repeats. The number of repeats has a stronger association with the hotspot phenotype than rs2914276, but the latter remains highly significant after accounting for the former. By sequencing the zinc-finger repeats from 55 Icelandic chromosomes covering the 12 to 15 repeat spectrum, a variant rs6875787 leading to an amino-acid change in the sixth zinc finger, also noted previously[4], was seen (Supplementary Fig. 4a). Additional sequencing and further investigations (Supplementary Information) showed that the minor allele of rs6875787 is in about 5.3% of the chromosomes with 13 repeats, and confirmed a previous suggestive finding[4] that it lowers the fraction of recombinations in hotspots. However, the effect is only about one-tenth that of the 14 or 15 repeats. Thus it could be that many polymorphisms in the *PRDM9* locus affect hotspots, but the repeat polymorphism alone captures most (>90%) of the association currently observed between variations in *PRDM9* and the hotspot phenotype. Because the 14 and 15 repeats do not behave that differently, we collapsed them into a single 14/15 allele with a frequency of about 3.9%. We estimated that the differences between the 14/15 and 12/13 repeats alone can account for 60% and 44% of the total systematic component of the hotspot phenotype for males and females, respectively (Supplementary Information). The 16 repeat found in the Hutterites[4] was not observed in the Icelandic samples examined, but the described differences between the 14/15 and the 12/13 repeats are novel.

Because few parents carry more than one 14/15 allele, we grouped heterozygous and homozygous carriers together (351 males and 429 females, contributing 502 and 612 meioses, respectively) to construct two sex-specific carrier recombination maps. The same was done for non-carriers: that is, those with only 12 or 13 repeats. Although carriers have fewer recombinations in hotspots, they remain in substantial excess of the genome average (average SRR = 13.1 and 11.4 for male and female carriers, respectively, compared with an average SRR of 19.0 and 16.9, respectively, for non-carriers). Moreover, the binding motif corresponding to 13 repeats is associated with increased recombination rate and hotspots in carriers and non-carriers, although the effect is slightly stronger for non-carriers (Supplementary Information and Supplementary Table 5). The binding motif predicted for 14 repeats is also associated with increased recombination rate for both sets of individuals, but here the effect is stronger for carriers. In addition to motif intensities at a bin itself, motif intensities at nearby bins also appear to have an effect (Supplementary Table 6). However, the magnitudes of all these correlations are low, and the motifs alone provide very modest power for predicting hotspots.

For the 10-kb bins, the correlation between the CEU and YRI maps is 0.716 (Supplementary Table 3). The correlation between our overall sex-averaged map is stronger with the COMBINED map (0.729) than with the CEU (0.700) and YRI (0.643) maps, which indicates that a substantial part of the difference between the CEU and YRI maps is noise. Nonetheless, by examining the variations of *PRDM9* in the HapMap YRI samples, we found zinc-finger repeat lengths of 12 to 15 and 17 to 19 (Supplementary Table 4 and Supplementary Fig. 4b). Grouping different repeat lengths into three composite alleles, the 12/13, 14/15 and 17/18/19 alleles have frequencies of 65.8%, 26.7% and 7.5%, respectively. The 14/15 allele is much rarer in the CEU samples, where only 13 and 14 repeats are found, at frequencies of 96.6% and 3.4%, respectively. We standardized all maps, including sex-averaged maps constructed separately for carriers and non-carriers of the 14/15 repeat (referred to as mapC and mapNC, respectively), in the same way as with our sex-specific maps. When regressing the difference between the YRI and CEU maps, on mapC and mapNC jointly, the coefficient of mapC was positive (0.089, $P < 10^{-100}$) and the coefficient of mapNC was negative (-0.307, $P < 10^{-300}$). When regressing the CEU and YRI maps separately on mapC and mapNC jointly, all coefficients were positive. For CEU, the coefficient of mapC was approximately 5.4% of the coefficient of mapNC, compared with 33.2% for YRI. Thus true differences between Europeans and Africans, explained by differences in frequencies of *PRDM9* variants, are identified.

There are 4,006 hotspot bins (10-kb bins with SRR >10) in our overall sex-averaged map, compared with 4,010 for the COMBINED map. The overlap is 2,139. Based on mapC and mapNC, 18.3% of the recombinations of carriers of the 14/15 repeat are in hotspots of our overall sex-averaged map, whereas the figure is 27.0% in non-carriers. Simulations that adjust for increased variation in maps estimated based on a reduced sample size show that, genome-wide, mapC is not smoother than mapNC, which suggests that the reduced recombination rate at hotspots defined by the overall map is compensated for by hotspots elsewhere. Among the 5,034 bins with an SRR greater than 10 in mapC, 1,380 and 371 have an SRR less than 3 and 1, respectively, in mapNC. A permutation test shows that 350 and 38 bins with such properties are expected by chance, suggesting that 1,030 of the 1,380 (74.6%), and 333 of the 371 (89.8%), are true hotspots specific to the 14/15 repeat carriers. Further support comes from examining the differences in SRR between the YRI and CEU maps, wherein 545 of the 1,380 (39.5%) and 150 of the 371 (40.2%) identified bins fall into the top fifth percentile of all bins studied.

In the genic regions, our sex-averaged map and the COMBINED map have an average SRR of 0.929 and 0.904, respectively. The difference is significant ($P = 3.1 \times 10^{-5}$) and is mainly accounted for by bins containing exons and intronic bins within 10 kb of an exon. One possibility

is that regions around exons are more likely to have been subject to natural selection, resulting in a lower number of recombinations detectable from LD and consequently leading to an underestimation of recombination rate in LD-based maps[7]. However, for regions (23,614 10-kb bins) that have been proposed as targets of selection by at least two of nine genome scans compiled by Akey[16], the difference between the COMBINED map and our map is opposite to that observed for regions around exons. Specifically, although both maps assign low recombination rates to these regions, the average SRR in the COMBINED map (0.650) is significantly higher ($P = 5.7 \times 10^{-9}$) than that in our sex-averaged map (0.593). Whether these differences are a result of some novel bias that affects estimates of LD-based recombination rate at regions under selection, or partly reflect properties exhibited by the statistical methods used to identify regions under selection that are currently poorly understood, warrants further investigation.

Polymorphisms at the *RNF* gene that influence total genome-wide recombination rates of males and females in opposite directions[15] have little impact on the fraction of recombinations in hotspots (Supplementary Information). Variations at the *PRDM9* gene influence recombination locations in a similar manner for both sexes, but have little effect on total genome-wide recombination rate. An inversion on chromosome 17 reported to associate with increased fertility and genome-wide recombination rate[17] also appears to increase the fraction of recombinations in hotspots, but the effect is limited to females ($P = 2.9 \times 10^{-5}$ and 0.49 for females and males, respectively) and is modest (Supplementary Information). These polymorphisms, together with the systematic regional and local differences in recombination rates between the sexes, provide a glimpse of nature's ingenuity in building diversity and flexibility into the system. The maps constructed in this study (available at http://www.decode.com/addendum) should serve as a valuable resource for genetics research for years to come.

METHODS SUMMARY

Subjects were 20,217 distinct individuals genotyped using various Illumina BeadChips and processed to determine parental origin. When searching for variants associated with the hot-spot phenotype, adding to SNPs used to determine recombinations, another 497,257 SNPs typed for a subset of the individuals were imputed into the others with methods used before. Variants at the *PRDM9* gene were similarly imputed. Imputations were not used for map construction. The LD-based maps were downloaded from https://mathgen.stats.ox.ac.uk/impute. The entire zinc-finger domain of the *PRDM9* gene was amplified with unique primers outside the repetitive region, avoiding homology to chromosome 16, for a total of 575 Icelanders, 30 CEU and YRI trios, and 74 Han Chinese in Beijing (CHB) and Japanese in Tokyo (JPT) samples. The amplified product was run on agarose gel to determine the number of zinc-finger repeats. For further analysis, 55 bands of different repeat lengths from Icelanders and 21 from YRI samples were isolated from the agarose gel, cloned and fully sequenced. Statistical tests used were mainly regression based, for example paired and unpaired *t*-tests, correlation tests and regressions. Genomic control[18] was used for the genome-wide association analysis with the hotspot phenotype. For map comparisons, to handle correlations among close-by bins, a procedure that permutes and flips chromosomes was used to calculate adjustment factors for the test statistics. Another randomization procedure that permutes individuals was used to estimate false discovery rates for sex-specific and new hotspots. See Supplementary Information for details.

1. McVean, G. A. *et al.* The fine-scale structure of recombination rate variation in the human genome. *Science* **304**, 581–584 (2004).
2. Myers, S., Bottolo, L., Freeman, C., McVean, G. & Donnelly, P. A fine-scale map of recombination rates and hotspots across the human genome. *Science* **310**, 321–324 (2005).
3. Myers, S., Freeman, C., Auton, A., Donnelly, P. & McVean, G. A common sequence motif associated with recombination hot spots and genome instability in humans. *Nature Genet.* **40**, 1124–1129 (2008).
4. Baudat, F. *et al.* PRDM9 is a major determinant of meiotic recombination hotspots in humans and mice. *Science* **327**, 836–840 (2010).
5. Myers, S. *et al.* Drive against hotspot motifs in primates implicates the *PRDM9* gene in meiotic recombination. *Science* **327**, 876–879 (2010).

6. Parvanov, E. D., Petkov, P. M. & Paigen, K. *Prdm9* controls activation of mammalian recombination hotspots. *Science* **327,** 835 (2010).
7. O'Reilly, P. F., Birney, E. & Balding, D. J. Confounding between recombination and selection, and the Ped/Pop method for detecting selection. *Genome Res.* **18,** 1304–1313 (2008).
8. Kong, A. *et al.* A high-resolution recombination map of the human genome. *Nature Genet.* **31,** 241–247 (2002).
9. Kong, A. *et al.* Detection of sharing by descent, long-range phasing and haplotype imputation. *Nature Genet.* **40,** 1068–1075 (2008).
10. Kong, A. *et al.* Parental origin of sequence variants associated with complex diseases. *Nature* **462,** 868–874 (2009).
11. Dempster, A. P., Laird, N. M. & Rubin, D. B. Maximum likelihood from incomplete data via the *EM* algorithm. *J. R. Stat. Soc. B* **39,** 1–38 (1977).
12. The International HapMap Consortium. A haplotype map of the human genome. *Nature* **437,** 1299–1320 (2005).
13. Lang, M. R., Patterson, L. B., Gordon, T. N., Johnson, S. L. & Parichy, D. M. Basonuclin-2 requirements for zebrafish adult pigment pattern development and female fertility. *PLoS Genet.* **5,** e1000744 (2009).
14. Broman, K. W., Murray, J. C., Sheffield, V. C., White, R. L. & Weber, J. L. Comprehensive human genetic maps: individual and sex-specific variation in recombination. *Am. J. Hum. Genet.* **63,** 861–869 (1998).
15. Kong, A. *et al.* Sequence variants in the RNF212 gene associate with genome-wide recombination rate. *Science* **319,** 1398–1401 (2008).
16. Akey, J. M. Constructing genomic maps of positive selection in humans: where do we go from here? *Genome Res.* **19,** 711–722 (2009).
17. Stefansson, H. *et al.* A common inversion under selection in Europeans. *Nature Genet.* **37,** 129–137 (2005).
18. Devlin, B. & Roeder, K. Genomic control for association studies. *Biometrics* **55,** 997–1004 (1999).

LETTER

On–line, voluntary control of human temporal lobe neurons

Moran Cerf[1,2,3], Nikhil Thiruvengadam[1,4], Florian Mormann[1,5], Alexander Kraskov[1], Rodrigo Quian Quiroga[1,6], Christof Koch[1,7]* & Itzhak Fried[2,8,9,10]*

Daily life continually confronts us with an exuberance of external, sensory stimuli competing with a rich stream of internal deliberations, plans and ruminations. The brain must select one or more of these for further processing. How this competition is resolved across multiple sensory and cognitive regions is not known; nor is it clear how internal thoughts and attention regulate this competition[1–4]. Recording from single neurons in patients implanted with intracranial electrodes for clinical reasons[5–9], here we demonstrate that humans can regulate the activity of their neurons in the medial temporal lobe (MTL) to alter the outcome of the contest between external images and their internal representation. Subjects looked at a hybrid superposition of two images representing familiar individuals, landmarks, objects or animals and had to enhance one image at the expense of the other, competing one. Simultaneously, the spiking activity of their MTL neurons in different subregions and hemispheres was decoded in real time to control the content of the hybrid. Subjects reliably regulated, often on the first trial, the firing rate of their neurons, increasing the rate of some while simultaneously decreasing the rate of others. They did so by focusing onto one image, which gradually became clearer on the computer screen in front of their eyes, and thereby overriding sensory input. On the basis of the firing of these MTL neurons, the dynamics of the competition between visual images in the subject's mind was visualized on an external display.

One can direct one's thoughts via external stimuli or internal imagination. Decades of single-neuron electrophysiology and functional brain imaging have revealed the neurophysiology of the visual pathway[1,2]. When images of familiar concepts are present on the retina, neurons in the human MTL encode these in an abstract, modality-independent[5] and invariant manner[6,7]. These neurons are activated when subjects view[6], imagine[8] or recall these concepts or episodes[9]. We are interested here in the extent to which the spiking activity of these neurons can be overridden by internal processes, in particular by object-based selective attention[10–12]. Unlike imagery, in which a subject imagines a single concept with closed eyes, we designed a competitive situation in which the subject attends to one of two visible superimposed images of familiar objects or individuals. In this situation, neurons representing the two superimposed pictures vie for dominance. By providing real-time feedback of the activity of these MTL neurons on an external display, we demonstrate that subjects control the firing activity of their neurons on single trials specifically and speedily. Our subjects thus use a brain–machine interface as a means of demonstrating attentional modulation in the MTL.

Twelve patients with pharmacologically intractable epilepsy who were implanted with intracranial electrodes to localize the seizure focus for possible surgical resection[13] participated. Subjects were instructed to play a game in which they controlled the display of two superimposed images via the firing activity of four MTL units in their brain (Fig. 1). In a prior screening session, in which we recorded activity from MTL regions that included the amygdala, entorhinal cortex, parahippocampal cortex and hippocampus, we identified four different units that responded selectively to four different images[6]. Each trial started with a 2-s display of one of these four images (the target). Subjects next saw an overlaid hybrid image consisting of the target and one of the three remaining images (the distractor), and were told to enhance the target ('fade in') by focusing their thoughts on it. The initial visibility of both was 50% and was adjusted every 100 ms by feeding the firing rates of four MTL neurons into a real-time decoder[14] that could change the visibility ratios until either the target was fully visible ('success'), the distractor was fully visible ('failure'), or until 10 s had passed ('timeout'; see Fig. 2, Supplementary Figs 3 and 4 and Supplementary Video). We considered subjects' 'trajectories' in the plane defined by time and by the transparency of the two images making up the hybrid (Fig. 2a).

The subjects manipulated the visibility of the hybrid image by any cognitive strategy of their choosing. Six out of 12 subjects reported in a follow-up interview that they focused on the concept represented by the target picture (most often a person) or closely allied associations. Subjects did not employ explicit motor strategies to control these four units (see Supplementary Information). Subjects participated without any prior training and with a striking success rate in a single session lasting around 30 min, reaching the target in 596 out of 864 trials (69.0%; 202 failures and 66 timeouts). Results were significant ($P < 0.001$, Wilcoxon rank-sum) for each subject (Fig. 3). Subjects successfully moved from the initial 50%/50% hybrid image to the target in their first trial in 59 out of 108 first trials (54.6%).

Testing the extent to which successful competition between the two units responsive to the two images depends on their being located in different hemispheres, in different regions within the same hemisphere or within the same region (Fig. 3b), revealed that 347 out of 496 trials involving inter-hemispheric competitions were successful (70.0%; 123 failures, 26 timeouts), 177 out of 256 intra-hemispheric but inter-regional competitions were successful (69.1%; 45 failures, 34 timeouts) and 72 out of 112 intra-regional competitions were successful (64.0%; 30 failures, 10 timeouts). There is no significant difference between these groups at the $P = 0.05$ level.

Every 'fading sequence' in each trial that every subject saw was based entirely on the spiking activity of a handful of neurons in the subject's brain. We recorded from a total of 851 units, of which 72 were visually responsive (see ref. 6 for definition of 'responsive') and were used for feedback. In light of the explicit cognitive strategies reported by subjects—enhancing the target and/or suppressing the distractor—the question arises whether successful fading was due to increasing firing

[1]Computation and Neural Systems, California Institute of Technology, Pasadena, California 91125, USA. [2]Department of Neurosurgery, University of California, Los Angeles, California 90095, USA. [3]Stern School of Business, New York University, New York, New York 10012, USA. [4]School of Computer Science, Carnegie Mellon University, Pittsburgh, Pennsylvania 15213, USA. [5]Department of Epileptology, University of Bonn, Bonn 53105, Germany. [6]Department of Engineering, University of Leicester, Leicester LE1 7RH, UK. [7]Department of Brain and Cognitive Engineering, Korea University, Seoul, 136-713, Korea. [8]Semel Institute for Neuroscience and Human Behavior, University of California, Los Angeles, California 90095, USA. [9]Functional Neurosurgery Unit, Tel-Aviv Medical Center, Tel-Aviv 64239, Israel. [10]Sackler Faculty of Medicine, Tel-Aviv University, Tel-Aviv 69978, Israel.
*These authors contributed equally to this work.
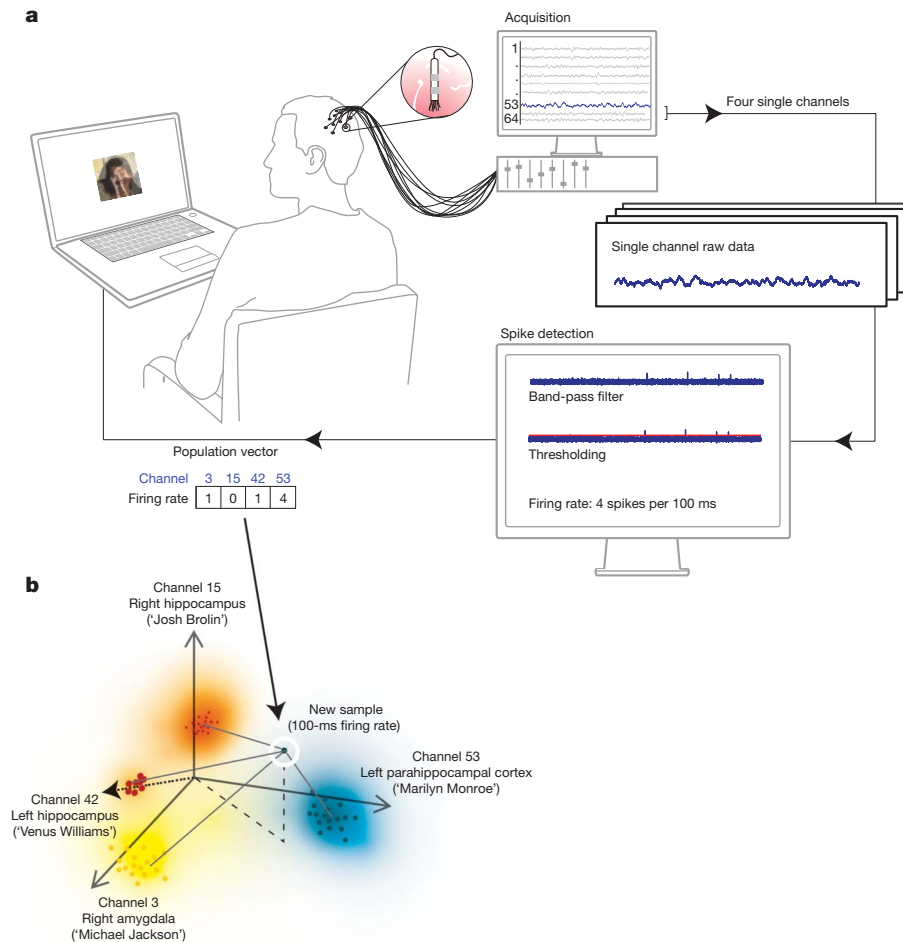
Figure 1 | Experimental set-up. a, Continuous voltage traces are recorded by 64 microelectrodes from the subject's medial temporal lobe. A four-dimensional vector, corresponding to the number of action potentials of four responsive units in the previous 100 ms, is sent to a decoding algorithm determining the composition of the hybrid seen by the subject with a total delay of less than 100 ms. **b**, The closest distance (weighted by the standard deviation) of this vector to the four clusters representing the four images is computed. If the 'winning' cluster represents the target or the distractor image, the visibility ratio of these two is adjusted accordingly.

of the unit the preferred stimulus of which was the target, to reducing the activity of the unit the preferred stimulus of which was the distractor or a combination of both. To answer this, we calculated firing rates in 100-ms bins in each trial for each unit. These rates were assigned to one of three categories labelled as follows. 'Towards target' meant the decoding process (based on the firing rate of all four units in this bin) enhanced the visibility of the target image, 'Away from target' meant decoding enhanced the distractor image and 'Stay' meant no change in visibility occurred (Supplementary Fig. 6). In the majority of successful trials (84.6%), the firing rate of the target-preferring unit was enhanced (3.72 standard deviations above baseline, $P < 10^{-4}$, t-test; Supplementary Fig. 7), simultaneously with suppression of the distractor-preferring unit (0.59 standard deviations below baseline, $P < 10^{-4}$, t-test). In 12.9% of successful trials only enhancement was seen, and in 1.1% only a reduction was seen. In the remaining trials, no significant deviation in baseline was detected. We observed no change in firing rates of the two units used for decoding, whose preferred stimuli were not part of the fading trial. Thus, successful fading was not caused by a generalized change in excitation or inhibition but by a targeted increase and decrease in the firing of specific populations of neurons. No long-lasting effect of feedback on the excitability of the MTL neurons was seen (see Supplementary Information).

To disentangle the effect of the retinal input from the instruction, we compared the activity of each unit in successful trials when the target

was the unit's preferred stimulus (target trials) with activity in successful trials when the target was the unit's non-preferred stimulus (distractor trials). This comparison was always done for the same retinal input, measured by the percentage of the visual hybrid allotted to the target (Fig. 4). We normalized each unit's response by its maximal firing rate over the entire experiment, and averaged over all trials for all subjects. For the same retinal input, the firing rate of neurons responding to the target pictures was much higher when subjects focused their attention on the target than when they focused on the distractor. The only difference was the mental state of the subject, following the instruction to suppress one or the other image.

To quantify the extent to which attention and other volitional processes dominate firing rates in the face of bottom-up sensory evoked responses, we devised a top-down control (TDC) index. TDC quantifies the level of control that subjects have over a specific unit and is the difference between the normalized firing rate when the subject attended the unit's preferred stimulus and the normalized rate when the subject attended the distractor image. That is, we subtracted the lower from the upper curve in Fig. 4a. Averaged over all 72 units, TDC equals 0.44 ± 0.28 (mean $\pm$ standard deviation), highly significantly different from zero. This was not true for failed trials (mean $P = 0.18$). If instead of subtracting the two curves the upper curve is divided by the lower one, a ratio of 6.17 ± 5.02 is obtained, highly significantly different from one. That is, the average unit fires more than six times as
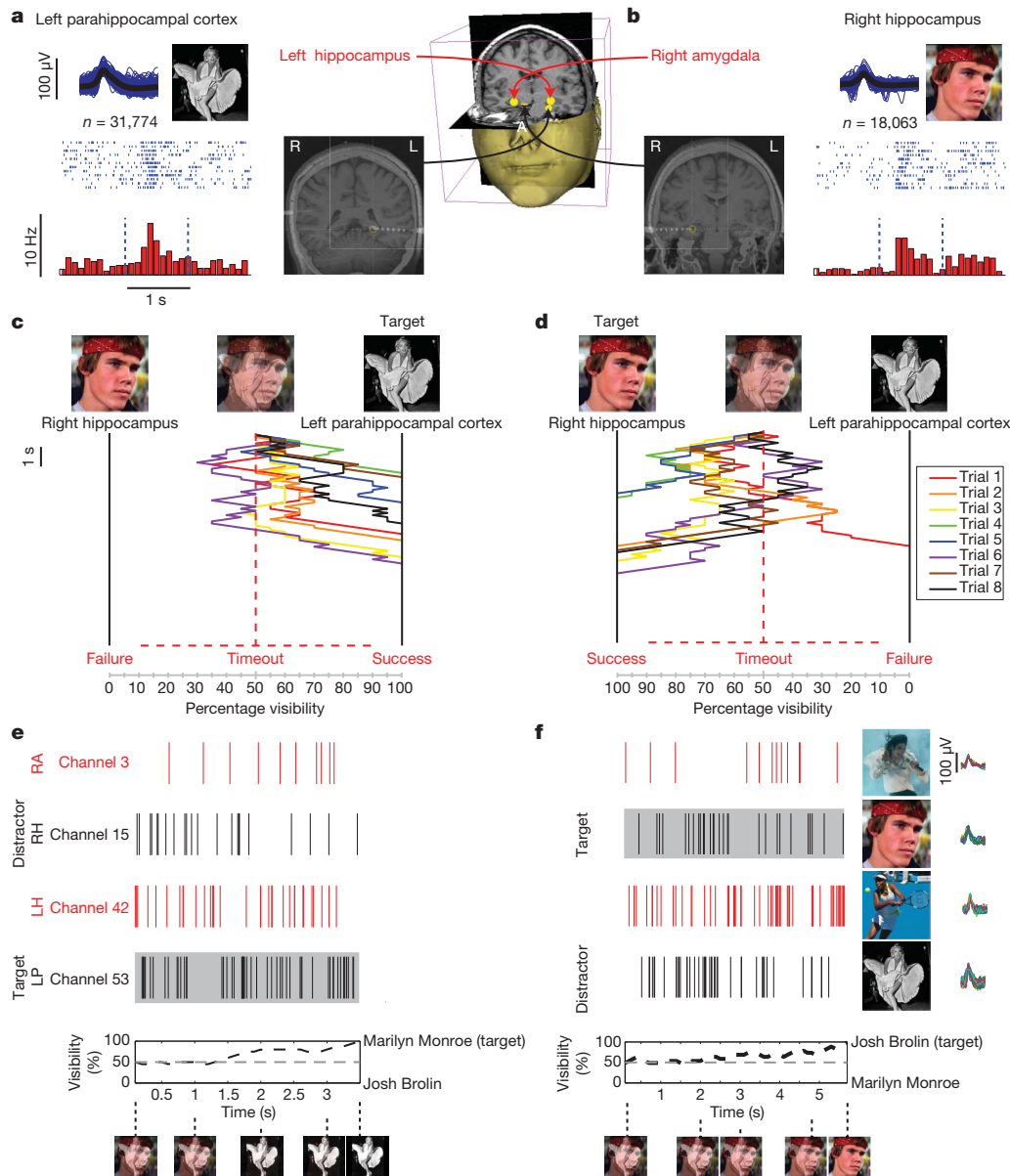
Figure 2 | Task performance and neuronal spiking. Two American actors, 'Josh Brolin' and 'Marilyn Monroe', constituted the preferred stimulus for two units. **a**, One multi-unit responded selectively to Monroe and was located in the left parahippocampal cortex. Below each illustration are the corresponding raster plots (twelve trials are ordered from top to bottom) and post-stimulus time histograms obtained during the control presentation. Vertical dashed lines indicate image onset (left) and offset (right), 1-s apart. Spike shapes are shown in blue, and the average spike shape in black. Below are the total number of spikes during the session. On the right is an illustration of the brain regions competing in these trials, and a fusion of the coronal CT and MRI scans taken after electrode implantation. Here, competing units were located in different hemispheres and regions. See Supplementary Video of the actual experiment. **c**, Time (running downwards for 10 s) versus percentage visibility of eight trials in which the subject had to fade a 50%/50% hybrid image into a pure Monroe image. The subject was able to do so all eight times, even though these were her first trials ever. **b**, **d**, When Brolin was the target, she succeeded seven out of eight times. All subjects show similar trends of controlled fading (Fig. 3). The hybrid image was controlled in real time by the spiking of four units selective to the image of Brolin, Monroe, Michael Jackson or Venus Williams. **e**, **f**, Spiking activity of all four units for one successful Monroe (**e**) and Brolin (**f**) trial. The spike shapes and the four images each unit is selective to are shown on the right. Below are the images as seen by the subject during the trial at different times. For another example, see Supplementary Figs 4 and 7. For copyright reasons, some of the original images were replaced in this and all subsequent figures by very similar ones (same subject, similar pose, similar colour and so on). The image of Josh Brolin is copyright The Goonies, Warner Bros. Inc. RA, right amygdala; RH, right hippocampus; LH, left hippocampus; LP, left parahippocampal cortex.

vigorously when the subject is attending to the unit's preferred image than when he/she is attending to the distractor. Excitation of the target unit, alongside inhibition of the distractor unit, occurs even in trials where the distractor is dominating the hybrid image, suggesting that the units are driven by voluntary cognitive processes capable of overriding distracting sensory input.

To control the extent to which successful 'fading in' was caused by the overall level of effort and attentional focus of the subject or by the instantaneous firing activity of the four units, we compared performance during normal feedback to that reached during sham feedback, when the image's visibility was, in fact, not guided by the subject's immediate neuronal activity but by activity from a previous trial (see Methods). Although subjects' level of effort and attention were the same as during real feedback, success dropped precipitously from 69.0% to 31.2% (33.7% failures and 35.1% timeouts; $\chi^2 = 69.9$, degrees of freedom = 2, $P < 10^{-4}$). Only two out of 12 subjects did better than chance during
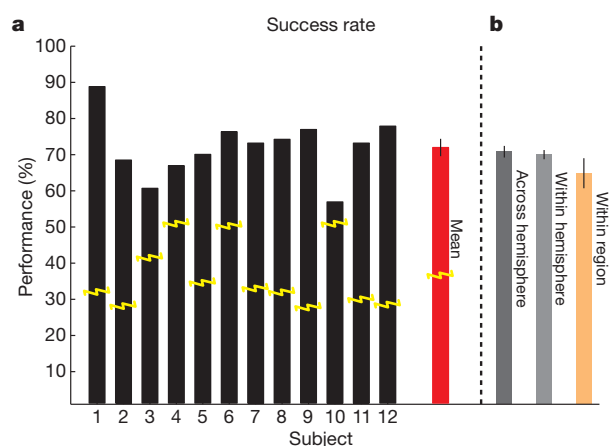
Figure 3 | Successful fading. a, Percentage of trials in which subjects successfully controlled the activity of four units and faded to the target image within 10 s. Yellow lines indicate chance performance—determined by bootstrapping 1,000 random trials for each subject ($P < 0.001$; Wilcoxon rank-sum). The red bar is the performance averaged over all 12 subjects. Error bars show the standard deviation. **b**, Percentage of successful trials of the entire data set in which the competition between the two units was across hemispheres, within the same hemisphere but in different regions, or within the same region. Error bars show standard deviations. Note that in **a**, performance is analysed across subjects, whereas in **b** it is analysed across eight trial fading sessions; hence, the means differ.

sham feedback ($P < 0.001$); the rest were not significant (P values: 0.15 ± 0.14). Furthermore, in contrast to the pattern observed with real feedback where subjects were able to successively delay failure over time (Supplementary Fig. 5), there was no such delay during sham feedback (see Supplementary Information). These findings support the notion that feedback from the four selective units controlling the composite

image were essential to carry out the task successfully, rather than the general cognitive efforts of the subject, exposure to the stimuli, or global changes in firing activity.

Our study creates a unique design within which to interrogate the mind's ability to influence the dominance of one of two stimuli by decoding the firing activity of four units deep inside the brain. The stronger the activity of the target-preferring unit and the weaker the activity of the distractor-preferring unit, relative to the two other units, the more visible the target became on the screen and the more opaque the superimposed distractor image became (and vice versa). Overall, subjects successfully 'faded-in' 69% of all trials. Cognitive processes voluntarily initiated by the subject, such as focusing on the target or suppressing the distractor, affected the firing activity of four units in different MTL regions, sometimes even across hemispheres (see Supplementary Information for list of all regions). The firing rate of these units generates a trajectory in a four-dimensional space. This was projected onto a one-dimensional walk along a line given by the competing representation of the target and the distractor image and visualized onto an external display. This path that subjects take may be analogous to the movement of rodents navigating in their physical environment using place fields[13].

The past decade has seen major strides in the development of brain–machine interfaces using single-neuron activity in the motor and parietal cortex of monkeys[15–18] and humans[19–22]. A unique aspect of the present study is the provision of feedback from regions traditionally linked to declarative memory processes. It is likely that the rapidity and specificity of feedback control of our subjects depends on explicit cognitive strategies directly matched to the capacity of these MTL neurons to represent abstract concepts in a highly specific yet invariant and explicit manner[5]. We previously estimated, using Bayesian reasoning, that any one specific concept is represented by up to one million MTL neurons, but probably by much less[23]. As our electrodes are sampling a handful of MTL neurons with predetermined selectivities[14], cognitive control
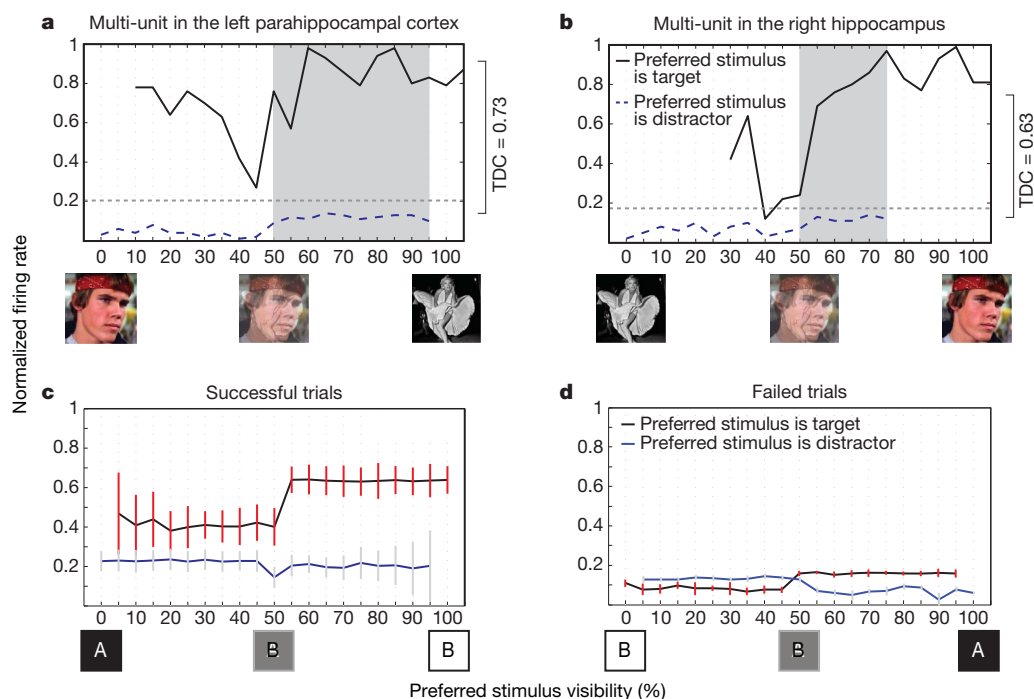
Figure 4 | Voluntary control at the single unit level. a, b, Normalized firing rates of the units in Fig. 2 as a function of visibility. We averaged the firing rates every 100 ms for every level of visibility for all successful trials where the target either was the unit's preferred (solid, black) or non-preferred stimulus (dashed, blue). Units fired significantly above baseline (grey dashed line) when the target was the preferred stimulus, and less than baseline when the target was the non-preferred stimulus. The TDC index is shown on the right. The shaded area reflects the bins used to calculate TDC. **c, d**, Averaging target and distractor trials across all subjects and all units for all successful fading trials reveals that the firing rate is significantly higher when the target is the preferred stimulus than in the competing situation, no matter what the visual input is. This is not true for failed trials (right). Red and dark grey vertical error bars are standard deviations. See Supplementary Fig. 8 for additional examples.

strategies such as object-based selective attention permit subjects to voluntarily, rapidly, and differentially up- and downregulate the firing activities of distinct groups of spatially interdigitated neurons to override competing retinal input. At least in the MTL, thought can override the reality of the sensory input. Our method offers a substrate for a high-level brain–machine interface using conscious thought processes.

METHODS SUMMARY

Subjects. Twelve patients with intractable epilepsy were implanted with depth electrodes to localize the epileptic focus for possible subsequent resection. The placement of all electrodes was determined exclusively by clinical criteria. All patients provided informed consent. All studies conformed to the guidelines of the Institutional Review Boards at UCLA and at Caltech.

Electrophysiology. Extracellular neural activity was acquired using 64 microwires implanted in various regions including the hippocampus, amygdala, parahippocampal cortex, and entorhinal cortex. Selected channels were band-pass filtered at 300–3,000 Hz, and a threshold was applied to detect spikes.

Experimental procedure. In a screening session, approximately 110 images of familiar persons, landmark buildings, animals, and objects were presented six times in random order for 1 s each. Four units were identified, each of which responded selectively to one of four different images. These four images were each presented 12 times to train a decoder. In a following fading experiment, each trial began with a 2-s presentation of the target. The subject then viewed a superposition of the target and one of the remaining three images, and was instructed to "continuously think of the concept represented by that image". Spike counts in 100-ms bins in the four selective units fully controlled the superposition on the screen in real time. At the end of the trial, acoustic feedback was given to the subject indicating success, failure or timeout after 10 s.

Data analysis. To evaluate each subject's performance, we used a bootstrapping technique—generating 1,000 random trials for each set of four units on the basis of their spiking activity and comparing their mean performance to that of the subject. Additionally, we analysed the activity of single and multi-units, compared against sham trials, compared unit activity across different regions, tested for changes in neuronal characteristics over time, and tested the level of control that subjects can exert over their neurons.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

1. Chalupa, L., Werner, J. & Barnstable, C. *The Visual Neurosciences* (MIT Press, 2004).
2. Thorpe, S. Single units and sensation: still just as relevant today. *Perception* **38**, 804–807 (2009).
3. Blake, D. A. R. (ed.) *Binocular Rivalry* (MIT Press, 2005).
4. Reynolds, J. & Chelazzi, L. Attentional modulation of visual processing. *Annu. Rev. Neurosci.* **27**, 611–648 (2004).
5. Quian Quiroga, R. *et al.* Explicit encoding of multimodal percepts by single neurons in the human brain. *Curr. Biol.* **19**, 1308–1313 (2009).
6. Quian Quiroga, R. *et al.* Invariant visual representation by single neurons in the human brain. *Nature* **435**, 1102–1107 (2005).
7. Földiák, P. Neural coding: non-local but explicit and conceptual. *Curr. Biol.* **19**, R904–R906 (2009).
8. Kreiman, G., Koch, C. & Fried, I. Imagery neurons in the human brain. *Nature* **408**, 357–361 (2000).
9. Gelbard-Sagiv, H. *et al.* Internally generated reactivation of single neurons in human hippocampus during free recall. *Science* **322**, 96–101 (2008).
10. Reddy, L., Kanwisher, N. & VanRullen, R. Attention and biased competition in multi-voxel object representations. *Proc. Natl Acad. Sci. USA* **106**, 21447–21452 (2009).
11. Desimone, R. & Duncan, J. Neural mechanisms of selective visual attention. *Annu. Rev. Neurosci.* **18**, 193–222 (1995).
12. Serences, J. *et al.* Control of object-based attention in human cortex. *Cereb. Cortex* **14**, 1346–1357 (2004).
13. Fried, I., MacDonald, K. & Wilson, C. Single neuron activity in human hippocampus and amygdala during recognition of faces and objects. *Neuron* **18**, 753–765 (1997).
14. Quian Quiroga, R. *et al.* Decoding visual inputs from multiple neurons in the human temporal lobe. *J. Neurophysiol.* **98**, 1997–2007 (2007).
15. Musallam, S. *et al.* Cognitive control signals for neural prosthetics. *Science* **305**, 258–262 (2004).
16. Wessberg, J. *et al.* Real-time prediction of hand trajectory by ensembles of cortical neurons in primates. *Nature* **408**, 361–365 (2000).
17. Velliste, M. *et al.* Cortical control of a prosthetic arm for self-feeding. *Nature* **453**, 1098–1101 (2008).
18. Moritz, C., Perlmutter, S. & Fetz, E. Direct control of paralysed muscles by cortical neurons. *Nature* **456**, 639–642 (2008).
19. Hochberg, L. *et al.* Neuronal ensemble control of prosthetic devices by a human with tetraplegia. *Nature* **442**, 164–171 (2006).
20. Kim, S. *et al.* Neural control of computer cursor velocity by decoding motor cortical spiking activity in humans with tetraplegia. *J. Neural Eng.* **5**, 455–476 (2008).
21. Kennedy, P. *et al.* Direct control of a computer from the human central nervous system. *IEEE Trans. Rehabil. Eng.* **8**, 198–202 (2000).
22. Guenther, F. *et al.* A wireless brain-machine interface for real-time speech synthesis. *PLoS ONE* **4**, e8218 (2009).
23. Waydo, S. *et al.* Sparse representation in the human medial temporal lobe. *J. Neurosci.* **26**, 10232–10234 (2006).

Author Contributions M.C., F.M., R.Q.Q., C.K. and I.F. designed the experiment; M.C. performed the experiments; I.F. performed the surgeries; M.C. and N.T. analysed the data; M.C., C.K. and I.F. wrote the manuscript. All authors discussed the data and the analysis methods and contributed to the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to M.C. (moran@klab.caltech.edu), C.K. (koch@klab.caltech.edu) or I.F. (ifried@mednet.ucla.edu).

METHODS

Subjects. Twelve patients participated in the study. Patients had pharmacologically intractable epilepsy and had been implanted with depth electrodes to localize the epileptic focus for possible subsequent resection. For each patient, the placement of the depth electrodes, in combination with microwires, was determined exclusively by clinical criteria[13]. All patients provided informed consent. All studies conformed to the guidelines of the Medical and Human subjects Institutional Review Boards at UCLA and the California Institute of Technology.

Screening. An initial morning screening session was recorded, during which approximately 110 images of familiar persons, landmark buildings, animals, and objects were presented six times in random order for 1 s each, after which each subject was asked to indicate with a button press whether the image contained a person or not. A standard set of such images was complemented by images chosen after an interview with the subject that determined which celebrities, landmarks, animals and objects the subject might be most familiar with. This approximately 30-min-long session—110 images × 6 repetitions × (1 s + reaction time)—was evaluated off-line to determine which of the 110 images elicited a response in at least one of 64 recorded channels, based on the criteria outlined in ref. 6. This involves measuring the median firing rate during the 300–1,000 ms after image onset across the six repetitions and comparing it to the baseline activity of the channel from 1,000–300 ms before image onset. Stimuli with median firing rates five standard deviations above baseline were considered selective.

From the group of selective units we chose four, based on their selectivity. The general guidelines for selection were: (1) to choose units from different brain regions so as to allow for competition between regions, (2) to select units that had similar characteristics in terms of latency and duration of the response within the 1 s the selective image is onscreen, and (3) to choose units for which the difference between firing rate during presentation and baseline was particularly clear. This selection was done by eye and was not quantitative.

Control presentations. The fading paradigm began with a short control presentations session—a presentation of the four selected images in random order, 12 repetitions at 1 s each—in a manner exactly replicating the set-up of the earlier screening session (see Supplementary Fig. 4 for results of the first control presentation for four units of one subject). The median firing activity over these 48 presentations between 1,000–300 ms before image onset determined the baseline firing rate for that unit for further statistical comparisons. The data from the control presentation procedure allowed for the set-up of a population-vector-based decoder.

We repeated the control presentation twice during each experiment—between the feedback blocks and at the end of the experiment, to verify that the neurons were still responsive for the stimuli used (Supplementary Fig. 1).

Fading. The following main fading experiment consisted of blocks of 32 trials each: eight for each of the four stimuli, shown in random order. Each trial began with a 2-s presentation of the target image. Subsequently, the subject viewed a superposition of the target image and one of the remaining three images (these two images were paired for the entire block). The hybrid image (H) was constructed from the target (T) and distractor image (D) by:

$$H = \alpha T + (1 - \alpha)D$$

where $\alpha \in [0, 1]$ corresponds to the trajectory in the images space—starting at 0.5 and changing in steps of 0.05 every 100 ms, ending either at 0 or 1 (see Supplementary Fig. 1 for illustration). α was controlled by the decoder, that is, ultimately by four units in the subject's brain.

The subject was instructed to enhance the target image from the hybrid image on the screen by "continuously thinking of the concept represented by that image". The subject was not directed in any further manner on what cognitive strategy to use—such as imagining that particular image or focusing on an aspect of the image—but was encouraged to explore the vast area of thoughts which might elicit a response. At the end of the trial, acoustic feedback was given to the subject indicating success, failure or timeout. The latter occurred after 10 s.

In each fading block (32 trials), two of the four images (say, A and B, together having 16 trials—eight trials with A as the target and eight with B as the target) received sham feedback, which did not reflect the neuronal activity during that trial. There was no overt difference between true and sham feedback trials. To achieve balanced exposure, any sham trial was a direct repetition of one prior real trial. For example, for a sham trial where image A was the target, the subject saw a hybrid image of A and B but the course of changes in each image's visibility was in fact based on the neuronal activity of a different previous trial (say, a trial with image C as the target and D as the distractor).

Decoding. Data from four selected channels (microwires) were read, and spikes were detected in real time for every 100-ms interval during the control presentation.

Each 1-s image presentation in the control presentation (four images × 12 repetitions) was broken into ten 100-ms bins. We used spikes from the seven bins from 300 ms to 1,000 ms following image onset for the analysis because these included the most relevant data for decoding[14]. The total numbers of spikes for each 100-ms bin formed clusters in a four-dimensional space representing the activity of the four units for each image. Thus, for 12 (repetitions) × 4 (images) × 7 (bins) we obtained a 336 (cluster) by 4 (channels) matrix corresponding to the firing rate during each image presentation for all 100-ms bins.

During fading, the firing rates from the four channels gave rise to a population vector that was used to associate the corresponding 100-ms bin to one of the four images. The population vector was a point in four-dimensional space, and we used the Mahalanobis distance to determine which cluster the point was closest to. The Mahalanobis distance was chosen as the distance measure because it is a fast and linear distance calculation measure that takes into account the shape of the cluster. Previous data showed that cluster variability is significant for our data[14], so taking the standard deviation of the cluster into account yielded better decoding.

The distance D from each of the four clusters is calculated as:

$$D = (\mathbf{x} - \bar{S}) \times \text{COV}(S)^{-1} \times (\mathbf{x} - \bar{S})^{\text{T}}$$

where $x = (x_1, x_2, x_3, x_4)$ is the new point in the four-dimensional space (corresponding to the firing rate of four units in the previous 100 ms). S is a 336 × 4 matrix of firing rates of four units during 100-ms bins in the control presentation when the subject was viewing one of four images (for example, columns 1:7 in the matrix correspond to seven 100-ms bins of the firing rates of the four channels while image A was on the screen, columns 8:14 correspond to activity while image C was on the screen, and so on) and $\bar{S}$ is the mean of S. $D = (d_1, d_2, d_3, d_4)$ where d_i corresponds to the distance from cluster i. COV is the covariance function.

The closest cluster was regarded as the concept the subject thought of. Notice that each trial consists of two concepts that, when decoded, directly influenced the visibility of the two associated images that make up the hybrid (annotated as A and B). Decoding of one of the other two concepts (annotated C and D) was interpreted as 'thinking of neither A nor B'. In any given 100 ms of each fading trial, there were three possible outcomes: (1) the sample was closest to the cluster representing image A, causing the transparency of image A to increase by 5% and the transparency of B to decrease by 5% in the hybrid image seen by the subject. That is, if the proportion of transparency of images A/B was 50%/50% in the previous 100 ms, it would change to 55%/45%. (2) The sample looked more like a sample in the cluster associated with image B, which would lead to a 5% fading in the direction of image B. (3) The outcome was that the sample looked more like images in clusters C or D. This did not result in any change in the hybrid image.

Any one trial could last as little as 1 s (ten consecutive steps from 50%/50% to 100%/0% or 0%/100%). A limit of 10 s was set for each trial, after which the trial was regarded as 'timeout' whatever the transparency of the two images. All the decoding parameters were based on the post-hoc decoding analysis done on a similar MTL population in ref. 14.

Set-up. The experiment was run on a 15-inch laptop computer with images of 160 × 160 pixels centred on the screen at a distance of about 50 cm from the subject (visual angle of each image of 5.30° × 5.36°). Data from the subject's brain was acquired using the Cheetah system (Neuralynx) at 28 kHz, from which it was sent to a server performing spikes detection. Four selected channels were band-pass filtered at 300–3,000 Hz, and a threshold was applied to detect spikes. This threshold was set before the experiment based on a 2-min recording from each channel while the subject was sitting still with eyes opened. Spike counts in the four channels, per 100-ms bin, were transferred via TCP/IP (transmission control protocol/internet protocol) to the experiment laptop computer where the data was used for the online manipulation of the hybrid image. The feedback operation took place in under 100 ms. The experiment was programmed using Matlab (Mathworks) and the Psychophysics toolbox (version 2.54), while the spikes detection proprietary software was written in C++ for efficiency and real-time analysis (code provided on the authors' website at http://www.klab.caltech.edu/~moran/fading).

Response characteristics. We analysed units from the hippocampus, amygdala, entorhinal cortex and parahippocampal cortex. We recorded from 64 microwires in each session. We identified a total of 133 units (68% multi-units and 32% single-units) that were responsive to at least one picture. Out of these responses we selected four in each of 18 sessions. Seven subjects ran one experiment (7 × 4 units), four subjects ran the experiment twice with two different sets of four units (4 × 4 × 2 units), and one subject had three sessions, each with a different set of four units (1 × 4 × 3 units) for a total of 72 units. Out of these responsive units, 58 multi-units and 14 single-units were used in the subsequent fading experiment

(see Supplementary Fig. 2 for a distribution of the units used, and Supplementary Fig. 9 for illustration of the regional competition and performance).

Responses were either positive (exhibiting an increase in the firing rate above baseline, where baseline was determined during the control presentation as described above), or negative (decreasing the firing rate). Excitation was determined using the following techniques developed in previous work[6], by considering the interval after trial onset for all successful trials, divided by the number of spikes. Inhibition was determined using the following four criteria: (1) the median number of spikes in the interval after trial onset for all successful trials, divided by the number of spikes, was at least two standard deviations below the baseline activity, (2) a paired t-test using $P = 0.05$ as significance level rejected the null hypothesis of equal means, (3) the median number of spikes during baseline was at least two, (4) the median difference between the number of spikes in the trial and the baseline interval was higher than the background activity of 95 randomly resampled responses (bootstrapping).

Single and multi-units. Spikes used in the analysis were not sorted (that is, clustered) by their shape, but were instead taken as multi-units. This was done to speed up the calculation because template matching of individual spikes on-line had to be sacrificed for the sake of real-time decoding with less than 100 ms delay. Post-hoc analysis of the theoretical performance we could expect had we clustered spikes suggests that it would have increased the performance by 8–10%; however, this is difficult to be sure of because any post-hoc analysis of our data are biased by the fact that we do not have the subjects' feedback to the improved visibility changes on the screen. A further improvement of the set-up would be an additional on-line sorting of spikes, which would lead to a decrease in noise.

Bootstrap testing of statistical significance for task performance. To compare the performance of individual subjects (as in Fig. 3) against chance level we used a bootstrapping technique—generating random trials of activity for each set of four units on the basis of their activity and comparing the mean performance of those to that of the subject. We set individual baselines in the following way: each subjects' sequence of 32 trials (8 trials $\times$ 4 images) was broken into individual 100-ms steps, such that the decoding result for each step was categorized as 'towards target', 'away from target', or 'stay'. For example, in the first trial (coloured red) on the left panel of Fig. 2c (where the target was Marilyn Monroe) the first six 100-ms steps were 'towards target', the seventh 100-ms step was 'towards distractor', the eighth was 'stay', and so on. Thus, each subject ended up having a total number of bins reflecting the proportions of steps he or she used during the course of the entire experiment. This proportion reflected the subject's own baseline chance of going in either direction (the subject in Fig. 2, for instance, had 389 steps where she went towards the target, 49 steps towards the distractor, and 18 'stay' steps altogether). Using these proportions as a priori probabilities, we generated 1,000 new 32-trial blocks. For each 100-ms step, we randomly generated a direction of movement based on the probabilities calculated for each subject, and then generated trials. For each block we calculated the performance and then compared the 1,000 realizations to the one the subject actually performed. If the subject's performance were based only on his/her personal biases (moving in a certain direction because of faster response onset by one unit, paying more attention repeatedly to one of the two competing concepts, and so on) then the random realizations should exhibit a similar performance. The subject's actual performance would be better than the random realizations only if the subject was able to use his or her moves accurately to manoeuvre the fading of the two images towards the target.

LETTER

The patterns and dynamics of genomic instability in metastatic pancreatic cancer

Peter J. Campbell[1,2]*, Shinichi Yachida[3]*, Laura J. Mudie[1], Philip J. Stephens[1], Erin D. Pleasance[1], Lucy A. Stebbings[1], Laura A. Morsberger[3], Calli Latimer[1], Stuart McLaren[1], Meng-Lay Lin[1], David J. McBride[1], Ignacio Varela[1], Serena A. Nik-Zainal[1], Catherine Leroy[1], Mingming Jia[1], Andrew Menzies[1], Adam P. Butler[1], Jon W. Teague[1], Constance A. Griffin[3], John Burton[1], Harold Swerdlow[1], Michael A. Quail[1], Michael R. Stratton[1,4], Christine Iacobuzio-Donahue[3] & P. Andrew Futreal[1]

Pancreatic cancer is an aggressive malignancy with a five-year mortality of 97–98%, usually due to widespread metastatic disease. Previous studies indicate that this disease has a complex genomic landscape, with frequent copy number changes and point mutations[1–5], but genomic rearrangements have not been characterized in detail. Despite the clinical importance of metastasis, there remain fundamental questions about the clonal structures of metastatic tumours[6,7], including phylogenetic relationships among metastases, the scale of ongoing parallel evolution in metastatic and primary sites[7], and how the tumour disseminates. Here we harness advances in DNA sequencing[8–12] to annotate genomic rearrangements in 13 patients with pancreatic cancer and explore clonal relationships among metastases. We find that pancreatic cancer acquires rearrangements indicative of telomere dysfunction and abnormal cell-cycle control, namely dysregulated G1-to-S-phase transition with intact G2–M checkpoint. These initiate amplification of cancer genes and occur predominantly in early cancer development rather than the later stages of the disease. Genomic instability frequently persists after cancer dissemination, resulting in ongoing, parallel and even convergent evolution among different metastases. We find evidence that there is genetic heterogeneity among metastasis-initiating cells, that seeding metastasis may require driver mutations beyond those required for primary tumours, and that phylogenetic trees across metastases show organ-specific branches. These data attest to the richness of genetic variation in cancer, brought about by the tandem forces of genomic instability and evolutionary selection.

We performed massively parallel paired-end sequencing to identify somatically acquired genomic rearrangements in 13 patients with pancreatic adenocarcinoma (Supplementary Table 1). For each sample, we generated 50–150-million paired sequences of 37 base pairs (bp) from 400–500-bp fragments of genomic DNA (Supplementary Figs 1 and 2). Putative rearrangements were screened by polymerase chain reaction (PCR) and capillary sequencing across the breakpoint, allowing annotation to base-pair resolution and distinction between germline and somatic rearrangements[13,14]. For three patients (patient IDs PD3644–PD3646), samples were early-passage cell lines from resected primary pancreatic tumours. For the other ten patients, multiple metastases were collected at autopsy. In seven of these (PD3637–PD3643), we performed paired-end sequencing on an early-passage cell line derived from a single metastasis per patient. In one patient (PD3826), we sequenced DNA from a bulky metastasis and in two patients (PD3827–PD3828), we separately sequenced three metastases per patient. Hereafter, we refer to lesions sequenced as 'index' metastases. For the ten patients with samples from multiple metastases, lesions not sequenced, as well as germline DNA, were genotyped by PCR for the presence or absence of each rearrangement.

We identified 381 somatically acquired and 177 germline rearrangements (Fig. 1a, Supplementary Tables 2 and 3), classified into 7 categories (Supplementary Table 4). The consequences of these rearrangements for protein-coding genes are discussed in Supplementary Results (see also Supplementary Figs 3 and 4 and Supplementary Tables 5 and 6). There was considerable inter-patient heterogeneity in patterns of genomic instability, with differences in numbers (3–65 per patient) and types of rearrangement ($P < 0.0001$; Fig. 1a). Genomic landscapes showed marked disparity within the cohort (Fig. 1b and Supplementary Fig. 5). For example, patient PD3640 had rearrangements evenly scattered across the genome, whereas 35/44 (80%) breakpoints from PD3641 involved chromosome 8. Intra-chromosomal rearrangements generally predominated over those between chromosomes, but in PD3646, an inter-crossing patchwork of joins among five chromosomes was the major feature in an otherwise quiet genome.

One sixth of rearrangements show a distinctive pattern we have termed 'fold-back inversions' (Fig. 1c). A copy number change is demarcated by read-pairs aligning close together but in inverted orientation. Thus, a genomic region is duplicated, but the two copies head away in opposite orientations from the breakpoint. We believe the most probable mechanism to be breakage-fusion-bridge cycles[15,16] (Supplementary Results and Supplementary Fig. 6). A double-stranded DNA break occurring in G0-1 phase is replicated during S phase, leading to two identical DNA ends. Repair pathways directly join these, leading to a fold-back inversion pattern at the junction and an unstable dicentric chromosome. We find that this form of genomic instability is an early event in the development of pancreatic cancer and, with marked similarities to data from mouse models[17], frequently underpins and initiates amplification of cancer genes (Supplementary Results and Supplementary Figs 7 and 8).

The distribution of rearrangements in pancreatic cancer is different to that observed in breast cancer[14] ($P < 0.0001$; Fig. 1d and Supplementary Fig. 9). In particular, deletions (22% versus 13%) and fold-back inversions (16% versus 2%) were more frequent in pancreatic cancer, whereas tandem duplications (8% versus 31%) and amplicon-related rearrangements (17% versus 28%) were less frequent.

Taken together, these data indicate that pancreatic cancer has a distinctive pattern of genomic instability. Breakage-fusion-bridge cycles predicate specific abnormalities of cell-cycle control, namely dysregulation of the G1-to-S transition and an intact G2–M checkpoint. Duplication of DNA breaks in S phase implies that repair was not required before DNA replication and end-to-end fusion of the duplicated breaks implies active G2–M surveillance. End-to-end chromosome fusions are often seen in association with telomere erosion and it may be that the double-strand DNA break initiating breakage-fusion-bridge repair results from telomere loss[5,17–20].

---

[1]Cancer Genome Project, Wellcome Trust Sanger Institute, Hinxton CB10 1SA, UK. [2]Department of Haematology, University of Cambridge, Cambridge CB2 2XY, UK. [3]Departments of Pathology and Oncology, Johns Hopkins Medical Institutions, Baltimore, Maryland 21287, USA. [4]Institute for Cancer Research, Sutton, Surrey SM2 5NG, UK.
*These authors contributed equally to this work.
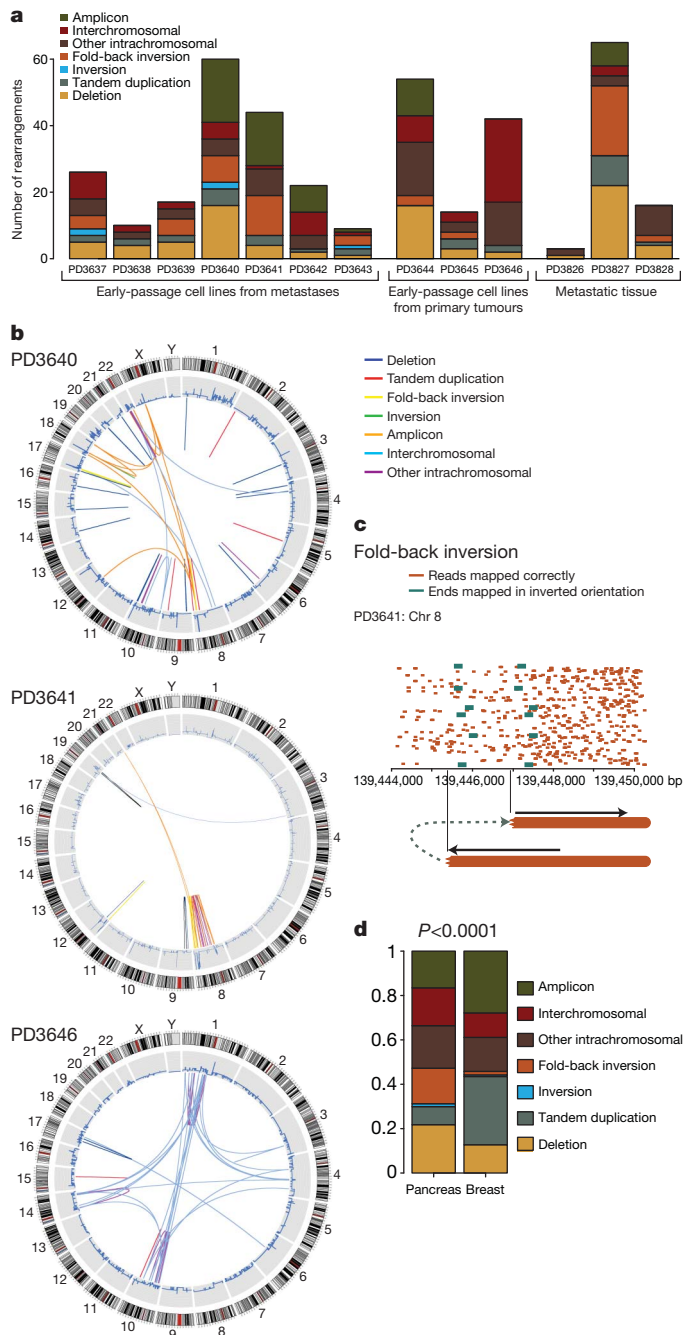
**Figure 1 | Patterns of somatically acquired genomic rearrangements in pancreatic cancer. a,** Histogram showing the distribution of the number and types of rearrangement observed in 13 patients with pancreatic cancer. **b,** Circle plots showing the genomic landscape of rearrangements in three representative samples. Chromosome ideograms are shown around the outer ring with copy number plots on the inner ring. Individual rearrangements are shown as arcs joining the two genomic loci, each coloured according to the type of rearrangement. **c,** Example of a so-called 'fold-back inversion'. Correctly mapping paired reads (orange) show much greater density on the right half of the figure than the left, indicating that the copy number is higher here. The change in copy number is demarcated by anomalously mapping paired reads (green), aligning ~2 kb apart on the genome and in inverted orientation. The only genomic structure that can explain this pattern is a rearrangement in which the abnormal chromosome is 'folded back' on itself leading to duplicated genomic segments in head-to-head (inverted) orientation. *x* axis, genomic position. Chr, chromosome. **d,** The distribution of types of rearrangement was significantly different between breast cancer and pancreatic cancer (*P* < 0.0001). *y* axis, proportion of all rearrangements.

To understand clonal relationships among metastases in pancreatic cancer, we genotyped 206 rearrangements across multiple lesions from ten patients (Figs 2–4, Supplementary Figs 10, 11 and Supplementary Table 7). Rearrangements followed three patterns: omnipresent across all lesions; partially shared by some but not all metastases; or unique to the index metastasis sequenced (Fig. 2a), with considerable inter-individual heterogeneity (Fig. 2b).

In comparison with other classes of rearrangement, fold-back inversions were significantly more likely to be found in all metastases from that patient (*P* = 0.003; Fig. 2c), implying that fold-back inversions occur early in cancer development, before tumour cells disseminate. Breakage-fusion-bridge cycles, resulting in fold-back inversions, are
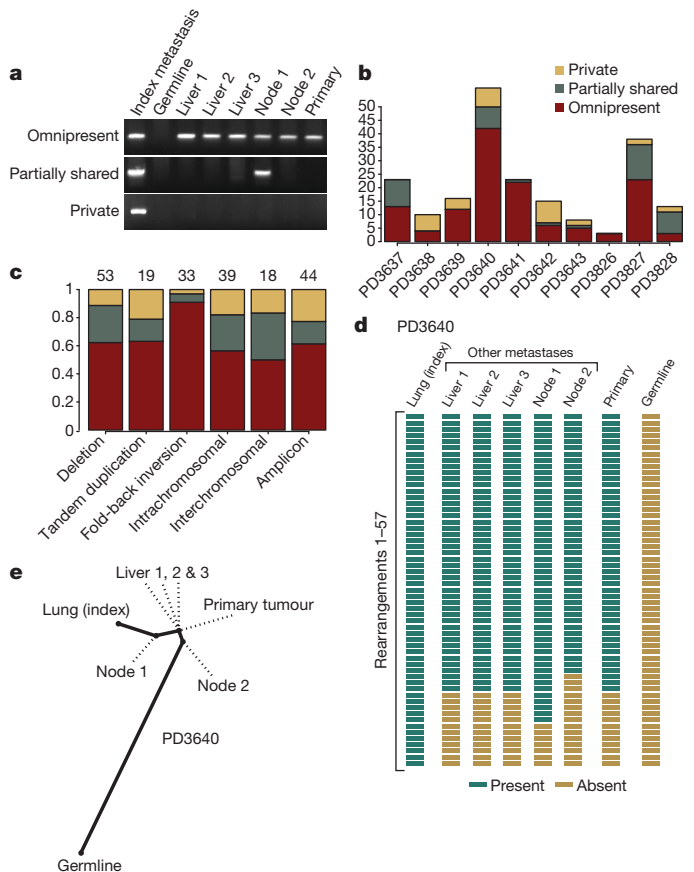
Figure 2 | Phylogenetic relationships of different metastases within a patient. a, PCR genotyping of three rearrangements across DNA from the index metastasis sequenced, other metastases from the same patient, the primary tumour and germline tissue. Somatic rearrangements may be present in all cancer samples but not the germline (omnipresent); present in some but not all metastases (partially shared); or present just in the index metastasis sequenced (private). **b,** Inter-individual differences in the proportions of rearrangements that are omnipresent across metastases, partially shared by some but not all lesions or are private to the index metastasis sequenced. **c,** Patterns across six broad categories of rearrangement in the proportions of variants that are omnipresent across metastases, partially shared by some but not all lesions or are private to the index metastasis sequenced. The numbers of rearrangements in each category are shown at the top. The difference in proportions between fold-back inversions and the other categories was statistically significant (*P* = 0.003). **d,** Genotyping of 57 rearrangements in PD3640 shows a coherent, nested structure, with 42 found in all metastases and the primary tumour, 7 found uniquely in the index tumour and 8 partially shared by some but not all metastases. **e,** The nested structure of rearrangements defines a phylogenetic tree of relationships among the metastases and primary tumour. The length of heavy black lines is proportional to the genetic distance between nodes. Dotted lines delineate the departure points of other, unsequenced lesions from the lineage between the germline genome and that of the index metastasis.
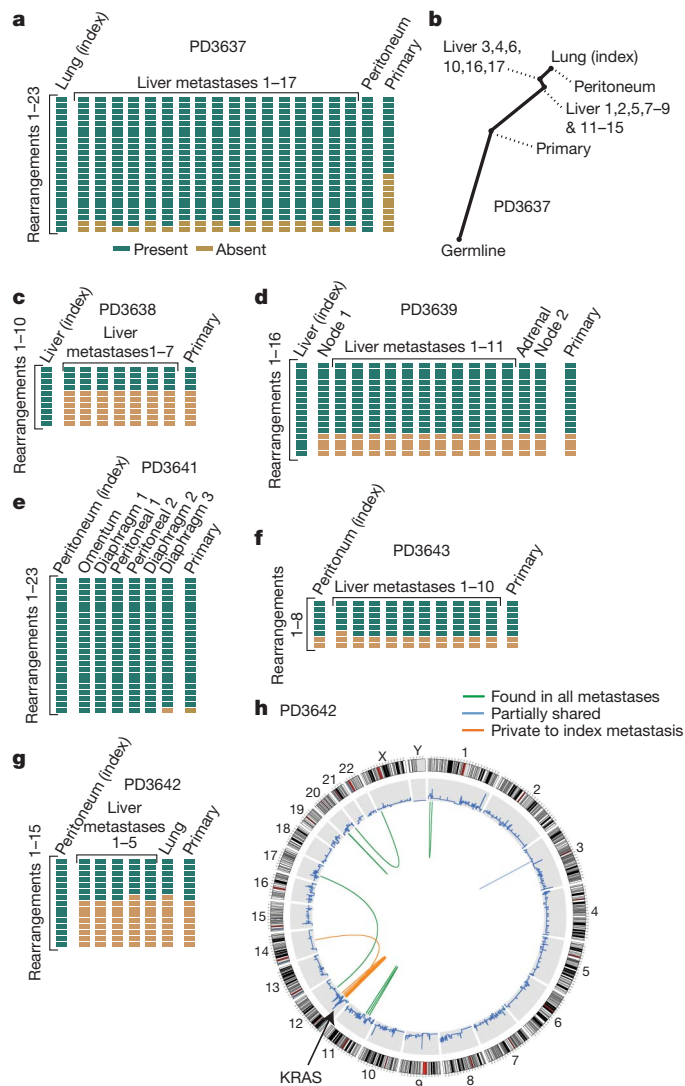
Figure 3 | Phylogenetic relationships among different metastases and the primary tumour. a, Results of PCR genotyping for 23 rearrangements across 19 metastases and the primary tumour from patient PD3637. **b**, Phylogenetic tree showing the relatedness of different metastases and the primary tumour. Note the early divergence of the primary tumour from all metastases. **c–g**, Genotyping results for PD3638 (**c**), PD3639 (**d**), PD3641 (**e**), PD3643 (**f**) and PD3642 (**g**). **h**, Circle plot showing that the rearrangements generating the amplicon of *KRAS* on chromosome 12 in PD3642 were only found in the index metastasis sequenced, and none of the other metastases or the primary tumour.

Figure 4 | Organ-specific signatures of metastasis. a, Results of PCR genotyping for 38 rearrangements across the 3 index metastases and 5 other metastases from patient PD3827. **b**, Overlapping out-of-frame deletions of exon 6 of *PARK2* were mutually exclusive to either the four lung metastases or the four abdominal metastases. The numbers above the gene refer to exon numbers and those below to the genomic position. **c**, A phylogenetic tree of relationships for metastases from patient PD3827, showing a clade of abdominal metastases and a further evolved clade of lung metastases. The length of heavy black lines is proportional to the genetic distance between nodes. Dotted lines delineate the departure points of other, unsequenced lesions from the lineage between the germline genome and that of the index metastasis. **d**, Results of PCR genotyping for PD3828. **e**, Phylogenetic tree of relationships for metastases from PD3828. **f**, Model for the clonal evolution of metastases derived from the patterns of phylogenetic relationships observed. Molecular time proceeds from left to right, and is associated with subclonal evolution and expansion within the developing primary tumour. Eventually a subclone within the primary tumour acquires the capacity to metastasize (pink), but this subclone continues to acquire genetic lesions (darkening shades of brown) such that different metastases may be founded from different clones. Within the developing metastases, clonal evolution continues, and these newly developed subclones can themselves seed tertiary metastases.

often initiated by telomere loss[5,18], whereas telomere attrition is not implicated in the pathogenesis of, for example, interstitial deletions or tandem duplications[21]. Telomerase, the gene that maintains telomere length, shows low expression during early pancreatic carcinogenesis before markedly increasing expression in the invasive tumour[5,18,22]. The genome-stabilizing effects of telomerase re-expression would therefore have more impact on reducing rates of fold-back inversion in advanced disease than other classes of rearrangement. In contrast, our data indicate that other types of rearrangement occur throughout the cancer life cycle, although the biological pathways underlying these forms of genomic instability remain unclear.

Subclonal evolution within tumours allows reconstruction of phylogenetic relationships[23]. Many rearrangements occur in the primary tumour before metastasis commences, and are therefore present in all metastases (Fig. 2b). However, in several patients, there is evidence for ongoing clonal evolution in the primary tumour among cells capable of initiating metastases. Three rearrangements in PD3640 are found in the

primary tumour and four metastases, but not the fifth (Fig. 2d, e), with a similar pattern in PD3642 (Fig. 3g). The most likely explanation is that two genetically distinct subclones of the primary independently seeded metastases. We cannot disprove that the discrepant metastasis lost the relevant rearrangements during clonal evolution, but the three events in PD3640 were on different chromosomes, making this unlikely. Importantly, these data indicate that metastasis is clonal, with individual deposits seeded by one or a few genetically similar cells, as described for prostate cancer[24].

We also find evidence of clonal evolution within metastases. Rearrangements private to the index lesion were found in seven out of ten patients. Most of these probably occurred in the developing

metastasis, although rearrangements acquired either in a subclone of the primary beneath the sensitivity of PCR or during *in vitro* passage[8] could give similar findings. Additionally, we found five rearrangements in PD3640 present in the index lesion and another metastasis but not in the primary tumour (Fig. 2a, d), with similar patterns in PD3637 (Fig. 3a) and PD3641 (Fig. 3e). These rearrangements might have arisen from clonal evolution in either a secondary metastasis that then itself seeded tertiary metastases or in a subclone of the primary that we have not sampled. Either way, there is considerable genetic heterogeneity among cells capable of initiating metastasis.

Whether metastasis requires mutations beyond those required to drive the primary tumour is controversial[25]. In PD3637, eight rearrangements were not found in the primary pancreatic tumour despite being present in all metastases (Fig. 3a, b and Supplementary Fig. 10). That all metastases are so phylogenetically distant from the primary tumour indicates that one or more driver mutations, which might either be among the eight rearrangements or among point mutations acquired contemporaneously, have conferred a selective advantage for metastatic spread. In published genomes from a matched breast cancer, brain metastasis and xenograft, there was similar enrichment in the metastasis and xenograft for 10–20 mutations at low prevalence in the primary, although driver mutations for metastasis could not be identified[8]. Taken together, these data imply the existence of a metastasis-promoting genomic signature in at least some patients.

We also find evidence for selection and adaptation within developing metastases after dissemination. For example, in a peritoneal metastasis from PD3642, *KRAS* is amplified to ~8–10 copies (Supplementary Fig. 7A). Because relevant sequencing reads all report the G12V mutation, amplification targeted the activating allele of *KRAS*. Remarkably, all rearrangements driving *KRAS* amplification were found only in the index metastasis and not in any other metastases or the primary (Fig. 3g, h). Within the index lesion, the rearrangements cause marked copy number changes, indicating that each is present in all tumour cells from that metastasis. This implies that rearrangements cumulatively amplifying mutant *KRAS* occurred early during establishment of the metastasis, driving successive waves of clonal expansion[26].

Little is known about whether metastases from a given organ system are more closely related to one another than to metastases from different organs. We therefore sequenced three metastases from two patients (Fig. 4). In PD3827, we identified two overlapping, out-of-frame deletions of exon 6 of *PARK2* (Fig. 4b). One was present in all four lung metastases but no abdominal deposits, whereas the other was carried by all four abdominal lesions but no lung deposits. Thus, the two deletions probably arose in separate clones, one of which founded the lung metastases and the other seeded the abdominal metastases. Similarly, in PD3828, lung metastases were on a separate branch of the phylogenetic tree from abdominal lesions (Fig. 4d).

In both patients, the lung lesions were further evolved than the abdominal metastases, and indeed, the additional rearrangements targeted cancer genes. Thus, similar to the *KRAS* amplicon in PD3642 described earlier, several of the lung-specific rearrangements might have conferred further selective advantage on that clone. In PD3828, eight rearrangements were restricted to lung metastases: these clustered around *MYC* and resulted in amplification not seen in abdominal metastases (Supplementary Fig. 11). Similarly, in PD3827, four out of twelve rearrangements restricted to lung metastases further amplified the *CCNE1* cancer gene (Supplementary Fig. 8B).

There are two explanations for organ-specific branches of phylogenetic trees. First, particular genotypes might drive metastasis to particular organs. The fact that lung metastases in these two patients were associated with additional driver mutations (amplification of *MYC* or *CCNE1*) indicates that tumour cells from subclones carrying these rearrangements were more likely to survive in the lung. Second, metastatic spread might be a stepwise process, occurring more readily within organ boundaries than between organs. These explanations are not mutually exclusive. Overcoming the barrier to colonizing a given organ might depend on a subclone of cancer cells acquiring particular adaptive changes, which, once established, can then disseminate through the organ with relative ease.

At first glance, the remarkable genetic diversity and adaptability of cancer under different selection pressures glimpsed here has ominous implications for our attempts to find curative therapies for metastatic disease. Nevertheless, for most patients studied here, more than half the rearrangements were found in all metastases and the primary tumour. The ability of studies such as this one to identify and understand these early mutations provides a route to the discovery of drug targets.

METHODS SUMMARY

Thirteen patients with pancreatic cancer were studied, with written informed consent for sample collection and analysis. Ten patients had multiple metastases collected at autopsy performed within 6 h of death, as described[27]. We also studied primary tumours collected from three patients undergoing resection with curative intent. Representative samples of primary carcinoma or metastases were minced with sterile blades, and the tissues gently pressed through a 45-μm mesh to disaggregate epithelial and stromal cells. For low-passage cell lines, filtered cells were resuspended into culture media and passaged up to five times to remove contaminating fibroblasts.

Protocols for massively parallel paired-end sequencing have been described in detail elsewhere[13,14]. Genomic DNA from the tumour samples was randomly fragmented, and fragments 400–500 bp in size selected by gel purification. Libraries were synthesized following our standard protocol, as described[28], and sequenced on a Genome Analyser II (Illumina) to give 37-bp reads from both ends of 50–150-million DNA fragments. In our experience, this identifies ~50–60% of rearrangements in a sample[13,14]. This level of genome coverage is insufficient to allow accurate identification of point mutations[11], but allows patterns of genomic rearrangement to be studied across several cancer samples without bias in size or type of rearrangement.

Sequencing data were aligned to the human reference genome (National Center for Biotechnology Information (NCBI) build 36) using the MAQ algorithm[29]. Clusters of anomalously mapping reads spanning putative rearrangements were identified informatically[13]. PCR across the breakpoint was performed in tumour and normal DNA, allowing rearrangements to be classified as somatically acquired, germline or artefactual. PCR products underwent capillary sequencing to annotate breakpoints to base-pair resolution. In ten patients, primers for somatic rearrangements were used to genotype by PCR all other metastases and, where available, the primary tumour from that patient. The sensitivity of PCR for detection of genomic rearrangements is at least 1/1,000 cells[30], considerably better than can be achieved for point mutations.

1. Jones, S. *et al.* Core signaling pathways in human pancreatic cancers revealed by global genomic analyses. *Science* **321,** 1801–1806 (2008).
2. Harada, T. *et al.* Genome-wide DNA copy number analysis in pancreatic cancer using high-density single nucleotide polymorphism arrays. *Oncogene* **27,** 1951–1960 (2008).
3. Fu, B., Luo, M., Lakkur, S., Lucito, R. & Iacobuzio-Donahue, C. A. Frequent genomic copy number gain and overexpression of GATA-6 in pancreatic carcinoma. *Cancer Biol. Ther.* **7,** 1593–1601 (2008).
4. Kimmelman, A. C. *et al.* Genomic alterations link Rho family of GTPases to the highly invasive phenotype of pancreas cancer. *Proc. Natl Acad. Sci. USA* **105,** 19372–19377 (2008).
5. Gisselsson, D. *et al.* Telomere dysfunction triggers extensive DNA fragmentation and evolution of complex chromosome abnormalities in human malignant tumors. *Proc. Natl. Acad. Sci. USA* **98,** 12683–12688 (2001).
6. Klein, C. A. Parallel progression of primary tumours and metastases. *Nature Rev. Cancer* **9,** 302–312 (2009).
7. Kuukasjarvi, T. *et al.* Genetic heterogeneity and clonal evolution underlying development of asynchronous metastasis in human breast cancer. *Cancer Res.* **57,** 1597–1604 (1997).
8. Ding, L. *et al.* Genome remodelling in a basal-like breast cancer metastasis and xenograft. *Nature* **464,** 999–1005 (2010).
9. Mardis, E. R. *et al.* Recurring mutations found by sequencing an acute myeloid leukemia genome. *N. Engl. J. Med.* **361,** 1058–1066 (2009).
10. Pleasance, E. D. *et al.* A comprehensive catalogue of somatic mutations from a human cancer genome. *Nature* **463,** 191–196 (2010).
11. Pleasance, E. D. *et al.* A small-cell lung cancer genome with complex signatures of tobacco exposure. *Nature* **463,** 184–190 (2010).
12. Shah, S. P. *et al.* Mutational evolution in a lobular breast tumour profiled at single nucleotide resolution. *Nature* **461,** 809–813 (2009).
13. Campbell, P. J. *et al.* Identification of somatically acquired rearrangements in cancer using genome-wide massively parallel paired-end sequencing. *Nature Genet.* **40,** 722–729 (2008).

14. Stephens, P. J. *et al.* Complex landscapes of somatic rearrangement in human breast cancer genomes. *Nature* **462,** 1005–1010 (2009).
15. McLintock, B. The stability of broken ends of chromosomes in *Zea mays. Genetics* **26,** 234–282 (1941).
16. Bignell, G. R. *et al.* Architectures of somatic genomic rearrangement in human cancer amplicons at sequence-level resolution. *Genome Res.* **17,** 1296–1303 (2007).
17. O'Hagan, R. C. *et al.* Telomere dysfunction provokes regional amplification and deletion in cancer genomes. *Cancer Cell* **2,** 149–155 (2002).
18. Bardeesy, N. & DePinho, R. A. Pancreatic cancer biology and genetics. *Nature Rev. Cancer* **2,** 897–909 (2002).
19. Maser, R. S. *et al.* Chromosomally unstable mouse tumours have genomic alterations similar to diverse human cancers. *Nature* **447,** 966–971 (2007).
20. Sahin, E. & Depinho, R. A. Linking functional decline of telomeres, mitochondria and stem cells during ageing. *Nature* **464,** 520–528 (2010).
21. Blow, J. J. & Gillespie, P. J. Replication licensing and cancer—a fatal entanglement? *Nature Rev. Cancer* **8,** 799–806 (2008).
22. Hashimoto Y, Murakami Y, Uemura K, *et al.* Telomere shortening and telomerase expression during multistage carcinogenesis of intraductal papillary mucinous neoplasms of the pancreas. *J. Gastrointest. Surg.* **12,** 17–29 (2008).
23. Campbell, P. J. *et al.* Subclonal phylogenetic structures in cancer revealed by ultra-deep sequencing. *Proc. Natl Acad. Sci. USA* **105,** 13081–13086 (2008).
24. Liu, W. *et al.* Copy number analysis indicates monoclonal origin of lethal metastatic prostate cancer. *Nature Med.* **15,** 559–565 (2009).
25. Nguyen, D. X. & Massague, J. Genetic determinants of cancer metastasis. *Nature Rev. Genet.* **8,** 341–352 (2007).
26. Klein, C. A. *et al.* Genetic heterogeneity of single disseminated tumour cells in minimal residual cancer. *Lancet* **360,** 683–689 (2002).
27. Embuscado, E. E. *et al.* Immortalizing the complexity of cancer metastasis: genetic features of lethal metastatic pancreatic cancer obtained from rapid autopsy. *Cancer Biol. Ther.* **4,** 548–554 (2005).
28. Quail, M. A. *et al.* A large genome center's improvements to the Illumina sequencing system. *Nature Methods* **5,** 1005–1010 (2008).
29. Li, H., Ruan, J. & Durbin, R. Mapping short DNA sequencing reads and calling variants using mapping quality scores. *Genome Res.* **18,** 1851–1858 (2008).
30. Flohr, T. *et al.* Minimal residual disease-directed risk stratification using real-time quantitative PCR analysis of immunoglobulin and T-cell receptor gene rearrangements in the international multicenter trial AIEOP-BFM ALL 2000 for childhood acute lymphoblastic leukemia. *Leukemia* **22,** 771–782 (2008).

Author Contributions P.J.C. undertook the analysis of the sequencing data assisted by P.J.S., E.D.P., L.A.S., M.-L.L., D.J.M., I.V., S.A.N.-Z., C.L., M.J., A.M., A.P.B. and J.W.T. Sample collection, processing, establishment of cell lines, DNA extraction and cytogenetic studies were performed by S.Y., L.A.M., C.A.G. and C.I.-D. PCR genotyping, capillary sequencing and downstream validation studies were performed by L.J.M. with assistance from C.L. and S.M. J.B., H.S. and M.A.Q. were responsible for generating libraries and running sequencers. P.J.C., S.Y., M.R.S., C.I.-D. and P.A.F. directed the research and wrote the manuscript, which all authors have approved.

LETTER

Distant metastasis occurs late during the genetic evolution of pancreatic cancer

Shinichi Yachida[1]*, Siân Jones[2]*, Ivana Bozic[3], Tibor Antal[3,4], Rebecca Leary[2], Baojin Fu[1], Mihoko Kamiyama[1], Ralph H. Hruban[1,5], James R. Eshleman[1], Martin A. Nowak[3], Victor E. Velculescu[2], Kenneth W. Kinzler[2], Bert Vogelstein[2] & Christine A. Iacobuzio-Donahue[1,5,6]

Metastasis, the dissemination and growth of neoplastic cells in an organ distinct from that in which they originated[1,2], is the most common cause of death in cancer patients. This is particularly true for pancreatic cancers, where most patients are diagnosed with metastatic disease and few show a sustained response to chemotherapy or radiation therapy[3]. Whether the dismal prognosis of patients with pancreatic cancer compared to patients with other types of cancer is a result of late diagnosis or early dissemination of disease to distant organs is not known. Here we rely on data generated by sequencing the genomes of seven pancreatic cancer metastases to evaluate the clonal relationships among primary and metastatic cancers. We find that clonal populations that give rise to distant metastases are represented within the primary carcinoma, but these clones are genetically evolved from the original parental, non-metastatic clone. Thus, genetic heterogeneity of metastases reflects that within the primary carcinoma. A quantitative analysis of the timing of the genetic evolution of pancreatic cancer was performed, indicating at least a decade between the occurrence of the initiating mutation and the birth of the parental, non-metastatic founder cell. At least five more years are required for the acquisition of metastatic ability and patients die an average of two years thereafter. These data provide novel insights into the genetic features underlying pancreatic cancer progression and define a broad time window of opportunity for early detection to prevent deaths from metastatic disease.

We performed rapid autopsies of seven individuals with end stage pancreatic cancer (Supplementary Table 1). In all patients, metastatic deposits were present within two or more anatomic sites in each patient, most often the liver, lung and peritoneum, as is typical for this form of neoplasia[4].

Low passage cell lines (six patients) or first passage xenografts (one patient) were created from one of the metastases present at each patient's autopsy. These samples comprised seven of the 24 pancreatic cancers which previously underwent whole exome sequencing and copy number analysis, as described in a mutational survey of the pancreatic cancer genome[5]. In this earlier study, a total of 426 somatic mutations in 388 different genes were identified among 220,884,033 base pairs (bp) sequenced in the seven index metastatic lesions, corresponding to an average of 61 mutations per index metastatic lesion (range 41–77). In all samples, the vast majority of mutations were represented by missense or silent single base substitutions (Supplementary Fig. 1 and Supplementary Table 2).

For each of the somatic mutations identified in the seven index metastasis lesions, we determined whether the same somatic mutation was present in anatomically distinct metastases harvested at autopsy from the same patients. We also determined whether these mutations

were present in the primary pancreatic tumours from which the metastases arose. A small number of these samples of interest were cell lines or xenografts, similar to the index lesions, whereas the majority were fresh-frozen tissues that contained admixed neoplastic, stromal, inflammatory, endothelial and normal epithelial cells (Fig. 1a). Each tissue sample was therefore microdissected to minimize contaminating non-neoplastic elements before purifying DNA.

Two categories of mutations were identified (Fig. 1b). The first and largest category corresponded to those mutations present in all samples from a given patient ('founder' mutations, mean of 64%, range 48–83% of all mutations per patient; Fig. 1b, example in Supplementary Fig. 2a). These data indicate that the majority of somatically acquired mutations present in pancreatic cancers occur before the development of metastatic lesions. All other mutations were characterized as 'progressor' mutations (mean of 36%, range 17–52% of all mutations per patient;
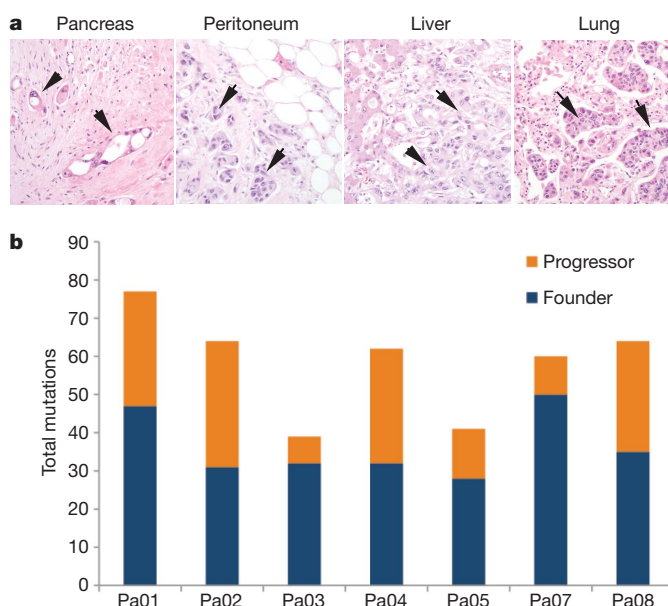
Figure 1 | Summary of somatic mutations in metastatic pancreatic cancers. a, Histopathology of primary infiltrating pancreatic cancer and metastatic pancreatic cancer to the peritoneum, liver and lung. In addition to infiltrating cancer cells in each lesion (arrows), non-neoplastic cell types are abundant. **b**, Total mutations representing parental clones (founder mutations), and clonal evolution (progressor mutations) within the primary carcinoma based on comparative lesion sequencing. Mutations common to all samples analysed were the most common category identified.

[1]Department of Pathology, The Sol Goldman Pancreatic Cancer Research Center, Johns Hopkins Medical Institutions, Baltimore, Maryland 21231, USA. [2]The Ludwig Center for Cancer Genetics and Therapeutics and The Howard Hughes Medical Institute at The Johns Hopkins Kimmel Cancer Center, Baltimore, Maryland 21231, USA. [3]Program for Evolutionary Dynamics, Department of Mathematics, Department of Organismic and Evolutionary Biology, Harvard University, Cambridge, Massachusetts 02138, USA. [4]School of Mathematics, University of Edinburgh, Edinburgh EH9 3JZ, UK. [5]Department of Oncology, The Sol Goldman Pancreatic Cancer Research Center, Johns Hopkins Medical Institutions, Baltimore, Maryland 21231, USA. [6]Department of Surgery, The Sol Goldman Pancreatic Cancer Research Center, Johns Hopkins Medical Institutions, Baltimore, Maryland 21231, USA.
*These authors contributed equally to this work.

©2010 Macmillan Publishers Limited. All rights reserved

Fig. 1b, example in Supplementary Fig. 2b). These mutations were present in one or more of the metastases examined, including the index metastasis, but not the parental clone.

These mutation types were used to classify the lesions that contained them into parental clones (containing only founder mutations) and subclones (containing both founder and progressor mutations). By definition, there could be only one parental clone in a patient, although there could be many different subclones. Parental clones tended to contain more deleterious mutations (nonsense, splice site or frameshift mutations) than subclones (12.6% of the mutations in the parental clones versus 8.1% of the mutations in subclones, Supplementary Table 2). The parental clones had already accumulated mutations in all driver genes (KRAS, TP53 and SMAD4) previously shown to drive pancreatic tumorigenesis[6]. Through combined analysis of high-density single nucleotide polymorphism (SNP) chip data on the index lesion (Supplementary Table 3) plus the sequencing data on all lesions (Supplementary Table 2) we found that the vast majority of homozygous mutations (51 mutations, representing 89% of all homozygous mutations) in the index lesion were already present in the parental clones. Homozygous mutations are characteristic of tumour suppressor genes such as SMAD4 and CDKN2A and often occur in association with chromosomal instability[7]. In sum, the parental clones harboured the majority of deleterious genetic alterations and chromosomal instability,

upon which were superimposed an accumulation of progressor mutations associated with clonal evolution and metastasis.

Evolutionary maps were constructed for each patient's carcinoma based on the patterns of somatic mutation and allelic losses and the locations of individual metastatic deposits (Fig. 2 and Supplementary Figs 3–8). These maps showed that, despite the presence of numerous founder mutations within the parental clones, the cells giving rise to the metastatic lesions had a large number of progressor mutations. For example, in Pa01 the parental clone contained 49 founder mutations, yet a clonal expansion marked by the presence of mutations in six additional genes was present in the lung and peritoneal metastases (Supplementary Fig. 3). Moreover, 22 more mutations were found in the liver metastasis. Note that all mutations in the metastatic lesions were clonal, that is, present in the great majority if not all neoplastic cells of the metastasis, as assessed by Sanger sequencing. Thus, these mutations were present in the cell that clonally expanded to become the metastasis. Similarly, large numbers of progressor mutations were generally observed in the metastases from each of the seven cases examined (Fig. 2 and Supplementary Figs 3–8).

To distinguish between the possibilities that clonal evolution occurred inside the primary cancer versus within secondary sites, we sectioned the primary tumours from two patients into numerous, three-dimensionally organized pieces (Fig. 2a, b) and examined the DNA
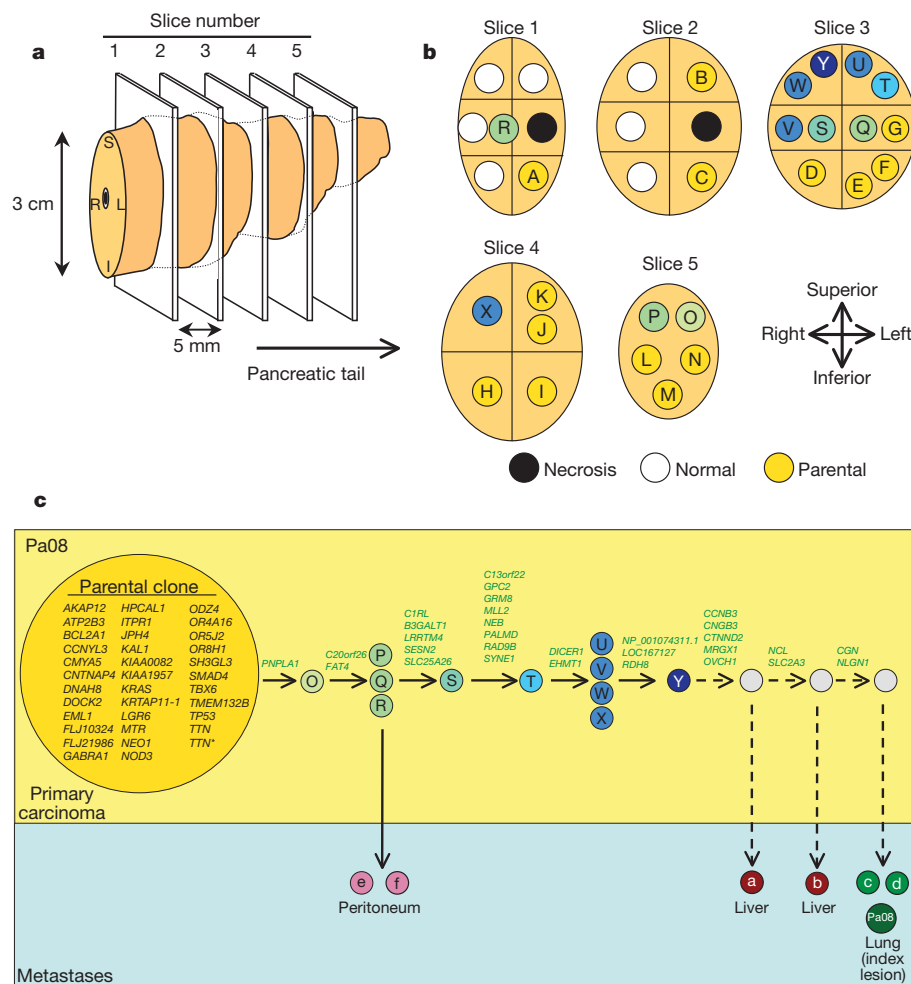
Figure 2 | Geographic mapping of metastatic clones within the primary carcinoma and proposed clonal evolution of Pa08. a, Illustration of the pancreatic specimen removed from Pa08 at rapid autopsy, and the planes of sectioning of the specimen. **b,** Mapping of the parental clone and subclones identified by comparative lesion sequencing within serial sections of the infiltrating pancreatic carcinoma. Metastatic subclones giving rise to liver and lung metastases are non-randomly located within slice 3, indicated by blue circles. These clones are both geographically and genetically distinct from clones giving rise to peritoneal metastases in this same patient, indicated in green. **c,** Proposed clonal evolution based on the sequencing data. In this model, after development of the parental clone, ongoing clonal evolution continues within the primary carcinoma (yellow rectangle), and these subclones seed metastases in distant sites. *Two mutations were found in the TTN gene.

from each piece for each of the founder and progressor mutations. In Patient Pa08, there were three progressor mutations present in two independent peritoneal metastases (defining one subclone) and 23, 25 or 27 additional progressor mutations present in liver and lung metastases (defining three additional subclones; Fig. 2c). Through the analysis of distinct regions of the primary tumour, it was clear that subclones giving rise to each of these metastases were present in the primary tumour. Moreover, these subclones were not small; from the size of the pieces (Fig. 2a) and the amounts of DNA recovered, each subclone must have contained in excess of 100 million cells. In addition, more than four different subclones, each containing a similarly large number of cells, could be identified through the analysis of other pieces of the same tumour. These subclones could be put into an ordered hierarchy establishing an evolutionary path for tumour progression (Fig. 2c). Analysis of multiple primary tumour pieces and metastatic lesions from patient Pa04 revealed a similar clonal evolution, with distinct, large subclones within the primary tumours giving rise to the various metastases (Supplementary Fig. 8).

To clarify further clonal evolution within the primary site, we attempted to correlate the mutation signatures representing the subclones of Pa08 (Fig. 2c) with the geographic location of the pieces of the primary tumour used to define them (Fig. 2a, b). Samples representative of the parental clone were located throughout the primary carcinoma. By contrast, samples representing subclones were non-randomly located in proximity to each other, within which the subclones specifically giving rise to peritoneal versus distant metastases were seen. Thus, we conclude that the genetic heterogeneity of metastases reflects heterogeneity already existing within the primary carcinoma, and that the primary carcinoma is a mixture of numerous subclones, each of which has independently expanded to constitute a large number of cells.

This data set could also be used to infer the timing of the development of the various stages of pancreatic tumour progression[8]. We assume that the tumour is initiated by a genetic event that confers a selective growth advantage to the cell that goes on to become the founder cell of the tumour. To estimate the timing, we first used Ki-67 labelling to determine the proliferation rate of seven samples of normal duct epithelium from surgically resected pancreata of individuals without pancreatic cancer as well as of each index metastasis. Ki-67-positive nuclei constituted an average of 0.4% of normal ductal cells, whereas an average of 16.3% of cancer cells within the index metastasis lesions were Ki-67-positive, consistent with prior estimates[9,10] (Supplementary Table 4). Based on these data plus that from sequencing of the index lesions, we derived estimates for three critical times in tumour evolution: T_1, the time between tumour initiation and the birth of the cell giving rise to the parental clone; T_2, the subsequent time required for the birth of the cell that gave rise to the index metastasis; and T_3, the time between the dissemination of this cell and the patients' death (Fig. 3). In other words, there is a time point, t_0, when the tumour was initiated, and a time point t_1 when a cell is born that has all mutations that exist in the parental clone. Similarly, there is a time point in tumour evolution, t_2, when a cell is born that has all the mutations that exist in the index metastasis. T_1 is given by $t_1 - t_0$ and T_2 is given by $t_2 - t_1$. If we denote t_3 as the time of patient's death, then $T_3 = t_3 - t_2$.

Using the mathematical model described in the Methods, we were able to conservatively estimate an average of 11.7 years from the initiation of tumorigenesis until the birth of the cell giving rise to the parental clone, an average of 6.8 years from then until the birth of the cell giving rise to the index lesion, and an average of 2.7 years from then until the patients' death (see Supplementary Discussion and Supplementary Table 5).

We show, for the first time, that primary pancreatic cancers contain a mix of geographically distinct subclones, each containing large numbers (hundreds of millions) of cells that are present within the primary tumour years before the metastases become clinically evident. The features of these metastatic subclones that promote metastasis formation have yet to be discerned, because no consistent genetic signature of metastatic subclones could be identified. We did identify several genes
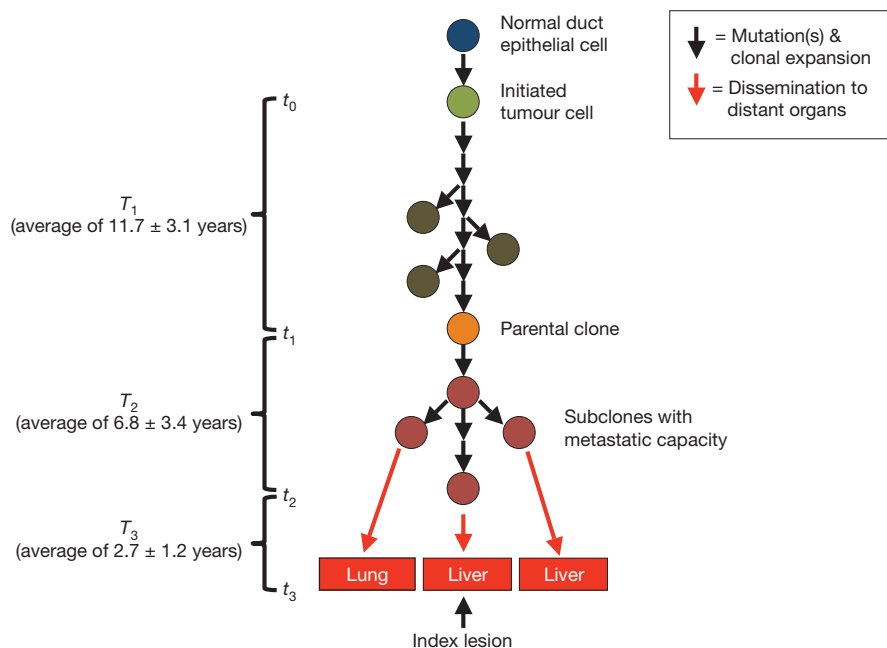
Figure 3 | Schema of the genetic evolution of pancreatic cancer.
Tumorigenesis begins with an initiating mutation in a normal cell that confers a selective growth advantage. Successive waves of clonal expansion occur in association with the acquisition of additional mutations, corresponding to the progression model of pancreatic intraepithelial neoplasia (PanIN) and time T_1. One founder cell within a PanIN lesion will seed the parental clone and hence initiate an infiltrating carcinoma (end of T_1 and beginning of T_2). Eventually, the cell that will give rise to the index lesion will appear (end of T_2 and beginning of T_3). Unfortunately, most patients are not diagnosed until well into time interval T_3 when cells of these metastatic subclones have already escaped the pancreas and started to grow within distant organs. The average time for intervals T_1, T_2 and T_3 for all seven patients is indicated in the parentheses at left (see also Supplementary Table 6).

that were mutated in one or more of the index metastatic lesions from these seven patients with Stage IV disease, but not in the primary pancreatic index lesions from 17 patients with Stage II disease (Supplementary Table 2). These genes include those that may have a role in invasive or metastatic ability through heterotypic cell adhesion (*CNTN5*), motility (*DOCK2*), proteolysis (*MEP1A*) and tyrosine phosphorylation (*LMTK2*). However, these mutations were not metastasis-specific *per se* as all but one were present in the matched primary carcinoma of those same seven patients, and there is no evidence that the mutations we observed endowed these genes with metastagenic activity. These data also do not reveal the selective pressures within the primary carcinoma that led to the formation of progressor mutations. In light of recent findings indicating that pancreatic cancers are poorly vascularized[11], one possibility is that intratumoural hypoxia creates a fertile microenvironment for the formation of additional mutations beyond that of the parental clone.

One of the major implications of these data is their implication for screening to prevent pancreatic cancer deaths. Quantitative analysis indicated a large window of opportunity for diagnosis while the disease was still in the curative stage—at least a decade. Our model also predicts an average of 6.8 years between the birth of the cell giving rise to the parental clone and the seeding of the index metastasis. Unfortunately, the great majority of patients are not diagnosed until the last 2 years of the entire tumorigenic process. The challenge is to detect these tumours during time T_1, or even after T_1 but before seeding of metastases. Advanced imaging methods, as well as blood tests to detect cancer-specific proteins, transcripts or genes[12], offer hope for such non-invasive early detection.

METHODS SUMMARY

Rapid autopsies were performed on seven individuals with Stage IV pancreatic cancer[13]. Genomic DNA was extracted from cell lines or xenografts established from one metastasis of each patient and used for exomic sequencing as described previously[5]. The Illumina Infinium II Whole Genome Genotyping Assay using the BeadChip platform was also used to analyse each sample at 1,072,820 (1M) SNP loci as described previously[5]. Samples of snap-frozen pancreatic cancer tissue were microdissected using a PALM MicroLaser System (Carl Zeiss MicroImaging) and DNA extracted using QIAamp DNA Micro Kits (Qiagen). Genomic DNA was quantified by calculating long interspersed nuclear elements (LINE) by real-time PCR. Whole genome amplification (WGA) was performed using 10 ng total template DNA and an illustra GenomiPhi V2 DNA Amplification Kit (GE Healthcare). Ki-67 immunolabelling (Clone MIB-1, Dako Cytomation) was performed on formalin-fixed, paraffin-embedded sections of normal pancreatic ducts and metastatic pancreatic cancer tissues for each patient using the Ventana Discovery staining system (Ventana Medical Systems), and this information was used to inform computational models of the timing of clonal evolution of each patient's pancreatic cancer (full details of these models are available in Full Methods).

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 11 May; accepted 15 September 2010.

1. Fidler, I. J. Critical determinants of metastasis. *Semin. Cancer Biol.* **12,** 89–96 (2002).
2. Nguyen, D. X., Bos, P. D. & Massague, J. Metastasis: from dissemination to organ-specific colonization. *Nature Rev. Cancer* **9,** 274–284 (2009).
3. Stathis, A. & Moore, M. J. Advanced pancreatic carcinoma: current treatment and future challenges. *Nature Rev. Clin. Oncol.* **7,** 163–172 (2010).
4. Yachida, S. & Iacobuzio-Donahue, C. A. The pathology and genetics of metastatic pancreatic cancer. *Arch. Pathol. Lab. Med.* **133,** 413–422 (2009).
5. Jones, S. *et al.* Core signaling pathways in human pancreatic cancers revealed by global genomic analyses. *Science* **321,** 1801–1806 (2008).
6. Maitra, A. & Hruban, R. H. pancreatic cancer. *Annu. Rev. Pathol.* **3,** 157–188 (2008).
7. Lengauer, C., Kinzler, K. W. & Vogelstein, B. Genetic instabilities in human cancers. *Nature* **396,** 643–649 (1998).
8. Jones, S. *et al.* Comparative lesion sequencing provides insights into tumor evolution. *Proc. Natl Acad. Sci. USA* **105,** 4283–4288 (2008).
9. Terada, T. *et al.* Cell proliferative activity in intraductal papillary-mucinous neoplasms and invasive ductal adenocarcinomas of the pancreas: an immunohistochemical study. *Arch. Pathol. Lab. Med.* **122,** 42–46 (1998).
10. Hisa, T. *et al.* Growth process of small pancreatic carcinoma: a case report with imaging observation for 22 months. *World J. Gastroenterol.* **14,** 1958–1960 (2008).
11. Olive, K. P. *et al.* Inhibition of Hedgehog signaling enhances delivery of chemotherapy in a mouse model of pancreatic cancer. *Science* **324,** 1457–1461 (2009).
12. Sidransky, D. Emerging molecular markers of cancer. *Nature Rev. Cancer* **2,** 210–219 (2002).
13. Embuscado, E. E. *et al.* Immortalizing the complexity of cancer metastasis: genetic features of lethal metastatic pancreatic cancer obtained from rapid autopsy. *Cancer Biol. Ther.* **4,** 548–554 (2005).
14. Fu, B. *et al.* Evaluation of GATA-4 and GATA-5 methylation profiles in human pancreatic cancers indicate promoter methylation patterns distinct from other human tumor types. *Cancer Biol. Ther.* **6,** 1546–1552 (2007).
15. Peiffer, D. A. *et al.* High-resolution genomic profiling of chromosomal aberrations using Infinium whole-genome genotyping. *Genome Res.* **16,** 1136–1148 (2006).
16. Bignell, G. R. *et al.* Signatures of mutation and selection in the cancer genome. *Nature* **463,** 893–898 (2010).
17. Sasaki, K. *et al.* Relationship between labeling indices of KI-67 and brdurd in human malignant tumors. *Cancer* **62,** 989–993 (1988).
18. Rew, D. A. & Wilson, G. D. Cell production rates in human tissues and tumours and their significance. Part II: clinical data. *Eur. J. Surg. Oncol.* **26,** 405–417 (2000).
19. Steel, G. G. *The Growth Kinetics of Tumours.* (Clarendon Press, Oxford, 1977).
20. Amikura, K., Kobari, M. & Matsuno, S. The time of occurrence of liver metastasis in carcinoma of the pancreas. *Int. J. Pancreatol.* **17,** 139–146 (1995).
21. Naumov, G. N. *et al.* A model of human tumor dormancy: an angiogenic switch from the nonangiogenic phenotype. *J. Natl. Cancer Inst.* **98,** 316–325 (2006).

Author Contributions Sample collection and processing was performed by C.A.I.-D., S.Y., B.F. and M.K. Microdissection, DNA extractions and whole genome amplification reactions were performed by S.Y. Sequencing was performed by S.J. Copy number analyses were performed by R.L. Computational models and estimates of clonal evolution were performed by I.B., T.A. and M.A.N.; C.A.I.-D., S.Y., S.J., R.H.H., J.R.E., M.A.N., I.B., T.A., V.E.V., K.W.K. and B.V. directed the research. C.A.I.-D., B.V., S.Y., I.B. and T.A. wrote the manuscript, which all authors have approved.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to C.A.I.-D. (ciacobu@jhmi.edu).

METHODS

Patients and tissue samples. Tissue samples from seven patients with pancreatic ductal adenocarcinoma were collected in association with the Gastrointestinal Cancer Rapid Medical Donation Program (GICRMDP). This programme was approved by the Johns Hopkins institutional review board and deemed in accordance with the Health Insurance Portability and Accountability Act. Details of the programme have been described in detail previously[13]. The tissue harvesting protocol consists of the following; after opening of the body cavity using standard techniques, the whole pancreas including the pancreatic cancer and each grossly identified metastasis were sampled using a sterile blade and forceps. The whole pancreas was sliced into $1 \times 1 \times 0.4$ cm sections for overnight fixation in 10% buffered-formalin, for freezing in Tissue-Tek OCT compound (Sakura Finetechnical) in liquid nitrogen and for snap-freezing in liquid nitrogen in 1.7 ml cryovials and storage at $-80\,°C$. Xenograft enriched or low passage cell lines were generated from the post mortem cancer tissues of these seven patients as described previously[13,14].

Laser capture microdissection (LCM). Frozen tissue sections of autopsy tissues were cut into 7 µm sections using a cryostat and embedded onto UV-treated PALM membrane slides (Carl Zeiss MicroImaging) and the slides were stored immediately at $-80\,°C$ until subsequent fixation. Tissue sections that underwent LCM were defrosted, fixed in 100% methanol for 3 min, and stained with toluidine blue before microdissection to remove contaminating stromal elements. Sections were dissected using a PALM MicroLaser System (Carl Zeiss MicroImaging). Dissected tissues were catapulted into adhesive caps. Generally, >20,000 cells were obtained from 5–10 serial sections by LCM to obtain sufficient quantity and quality of genomic DNA for subsequent amplification and sequencing.

Genomic DNA extraction and whole genome amplification. Genomic DNA from microdissected tissues was extracted using a QIAmp DNA Micro Kit (Qiagen) according to the manufacturer's protocol. Genomic DNA was quantified by calculating long interspersed nuclear elements (LINE) by real-time PCR. The LINE primer set 5′-AAAGCCGCTCAACTACATGG-3′ (forward) and 5′-TGCTTTGA ATGCAGAGAG-3′ (reverse) was designed. The real-time PCR conditions were 95 °C for 10 min; 40 cycles of 94 °C for 10 s, 58 °C for 15 s and 70 °C for 30 s. PCR was carried out using Platinum SYBR Green qPCR SuperMix-UDG (Invitrogen). To minimize sequencing bias from using low-copy starting templates, only samples for which the measured concentration by LINE assay was ≥ 3.3 ng µl^{-1} (1,000 genome equivalents) were used as a starting template for whole genome amplification (WGA). WGA was performed using 10 ng total template DNA and an illustra GenomiPhi V2 DNA Amplification Kit (GE Healthcare), following the manufacturer's protocol. WGA products were purified using a Microspin G-50 system (GE Healthcare). The purified WGA products were quantified by NanoDrop spectrophotometer (Thermo Fisher Scientific) and diluted to 20 ng µl^{-1} for sequencing analysis. Using these methods and quality controls, there was complete concordance in the mutational signatures obtained of cultured cell lines/ xenografts versus WGA materials prepared from their matched frozen tissues.

Sanger sequencing. PCR amplification and sequencing was performed using the conditions and primers described previously[5]. A small number of sequencing reactions failed (<2% of the total reactions) and these corresponding genes were not included in progression models or quantitative time estimates of clonal evolution.

Genotyping. The Illumina Infinium II Whole Genome Genotyping Assay using the BeadChip platform was used to analyse tumour samples at 1,072,820 (1M) SNP loci as previously described[5]. Briefly, all SNP positions were based on the hg18 (NCBI Build 36, March 2006) version of the human genome reference sequence. The genotyping assay begins with hybridization to a 50-nucleotide oligonucleotide, followed by a two-colour fluorescent single-base extension. Fluorescence intensity image files were processed using Illumina BeadStation software to provide normalized intensity values and allelic frequency for each SNP position. For each SNP, the normalized experimental intensity value (R) was compared to the intensity values for that SNP from a training set of normal samples and represented as a ratio (called the 'log R ratio') of $\log_2 (R_{experimental}/R_{training\ set})$. For each SNP, the normalized allele intensity ratio (theta) was used to estimate a quantitative allelic frequency value (called the 'B allele frequency') for that SNP[15]. Using Illumina BeadStudio software, log R ratio and B allele frequency values were plotted along chromosomal coordinates and examined visually. Regions of loss of heterozygosity (LOH) were identified as genomic regions >2 megabases (Mb) with consecutive homozygous genotype calls (B allele frequency near 0 or 1). Smaller (<2 Mb) regions of LOH were identified by requiring co-occurrence of decreased log R ratio scores in regions of consecutive homozygous genotype calls (B allele frequency near 0 or 1). Visual analysis of these data plotted along chromosomal coordinates was followed by manual analysis of the data for selected genes of interest.

Estimations of proliferation rates. To estimate the cell division rate, the Ki-67 labelling index (LI) in the proband lesion for each case was calculated. The Ki-67 LI on the pancreatic ducts in the histologically normal pancreas parenchyma was also

calculated. Normal pancreas was collected from two autopsied patients who died of causes other than pancreatic cancer and five patients who underwent distal pancreatectomy for a serous cystadenoma or an islet cell tumour at The Johns Hopkins Hospital. Paraffin blocks were cut into sections 4-µm thick for Ki-67 immunostaining with all staining processes from deparaffinization to counterstaining with haematoxylin being performed automatically with the Ventana Discovery staining system (Ventana Medical Systems). An anti-human Ki-67 mouse monoclonal antibody (Clone MIB-1, Dako Cytomation) was used. At least 12 randomly selected high-power fields containing a minimum of 2,000 cells were evaluated for each case, and the labelling index (LI) was calculated as the percentage of positive cell nuclei. Reactive small lymphocytes in each case were regarded as internal positive controls for Ki-67. Equal or more intense nuclear staining in comparison with the internal positive controls was considered to indicate positivity.

Modelling tumour evolution. Passenger mutations were defined as those unlikely to drive tumorigenesis. To be conservative, we considered passenger mutations as those not included as candidate cancer genes in a recent study based on whole exome sequencing of 24 pancreatic cancers[5]. As the great majority of mutations identified in cancers are believed to be passengers, the results of the model are not highly dependent on the model used to estimate the relatively small number of drivers[16].

Because passenger mutations are neutral and do not affect the evolution in any way, they are accumulated independently in each cell lineage. Following the lineage of the founder cell of the parental clone back in time, we can assume that it acquired a new neutral mutation with rate r at each cell division, with r being the product of the mutation rate per base pair per cell division and the number of base pairs sequenced. The accumulation of neutral mutations in a cell lineage can be well-described by a Poisson process with rate r per cell division. We are interested in the number of cell divisions in the single lineage between tumour initiation and birth of the founder cell of the parental clone during which N_1 passenger mutations accumulate. On the other hand, N_1 is also the number of mutations that are found in all tumour samples from one patient. Since we sequenced at least one sample from the primary tumour and at least three samples from different metastases from each patient, these specific N_1 mutations had to be present in the founder cells of all three metastases and in cells in the primary tumour. Thus there was a cell in the tumour that had these N_1 specific mutations for the first time, and that is, by definition, the founder cell of the parental clone. Since we can neglect the accumulation of mutations before the onset of the tumour, these N_1 mutations are accumulated along the single lineage from the tumour initiator cell to the founder cell of the parental clone. As the number of cell divisions between two subsequent mutations is distributed according to an exponential distribution with mean $1/r$, the required number of cell divisions is the sum of N_1 independent exponentially distributed random variables with mean $1/r$, and is distributed according to a Gamma distribution with shape parameter N_1 and scale parameter $1/r$. The mean of this distribution is N_1/r and the standard deviation is $\sqrt{N_1}/r$ (see Supplementary Table 6). Because the number of base pairs sequenced in the study is 31.7×10^6, and the mutation rate per base pair per generation is estimated at 5×10^{-10}, $r = 31.7 \times 10^6 \times 5 \times 10^{-10} \approx 0.016$ per generation[8].

Using our measurements of Ki-67 labelling index of the seven index lesions (average 16.3%), we were able to estimate the S-phase fraction of cells in the seven index lesions (average LI = 9.5%)[17]. Assuming a median value for the S-phase duration in human tissues and tumours, T_s, of 10 h (ref. 18) and using the formula for the potential cell doubling time $T_{pot} = \lambda T_s/LI$, we get an estimate for T_{pot} of 3.5 days. Here λ is a correction factor for the nonlinear age distribution of cells through the cell cycle, which was assumed to be 0.8 (ref. 19). This estimate is consistent with the average cell doubling time in pancreatic cancer from ref. 20 of 2.3 days. We use this latter estimate in our analysis, as we believe it is more accurate for pancreatic cancer.

Our model works very well for estimating the number of cell divisions between discrete events in tumour evolution. In order to go from number of cell divisions to actual time we need to have an estimate for the average rate of cell division. The accuracy of our predictions regarding actual time therefore depends on the accuracy of that estimate. If we let T_{gen} denote the average time between subsequent cell divisions in a cell lineage, we arrive at the expression for time T_1:

$$T_1 = \frac{T_{gen}}{r}(N_1 \pm \sqrt{N_1}).$$

We therefore estimate the number of cell divisions, and hence the time T_1 between tumour initiation and birth of the founder cell of the parental clone, to be proportional to the number of passenger mutations, N_1, that the tumour acquired during that time. In our calculations, we use the estimate for cell doubling time in pancreatic cancer from the literature[20] as the value of T_{gen}.

T_2 is determined analogously, with N_2 defined as the number of passenger mutations present in the index lesion but not in the parental clone. T_3 is

determined from literature-based estimates of the tumour and cell doubling times, and the size of the index lesions at autopsy[20].

The median doubling time of pancreatic cancer metastases was reported as 56 days[20]. To estimate the age of the index metastasis, we used a two stage model. We estimated the tumour doubling time was equal to the cell doubling time (T_{gen}) until the tumour size reached 100 μm in diameter at which time angiogenesis is required[21]. Thereafter, we used the median doubling time described above.

LETTER

Structural basis of semaphorin–plexin signalling

Bert J. C. Janssen[1]*, Ross A. Robinson[1]*, Francesc Pérez-Brangulí[2], Christian H. Bell[1], Kevin J. Mitchell[2], Christian Siebold[1] & E. Yvonne Jones[1]

Cell-cell signalling of semaphorin ligands through interaction with plexin receptors is important for the homeostasis and morphogenesis of many tissues and is widely studied for its role in neural connectivity, cancer, cell migration and immune responses[1]. SEMA4D and Sema6A exemplify two diverse vertebrate, membrane-spanning semaphorin classes (4 and 6) that are capable of direct signalling through members of the two largest plexin classes, B and A, respectively[2,3]. In the absence of any structural information on the plexin ectodomain or its interaction with semaphorins the extracellular specificity and mechanism controlling plexin signalling has remained unresolved. Here we present crystal structures of cognate complexes of the semaphorin-binding regions of plexins B1 and A2 with semaphorin ectodomains (human PLXNB1$_{1-2}$–SEMA4D$_{ecto}$ and murine PlxnA2$_{1-4}$–Sema6A$_{ecto}$), plus unliganded structures of PlxnA2$_{1-4}$ and Sema6A$_{ecto}$. These structures, together with biophysical and cellular assays of wild-type and mutant proteins, reveal that semaphorin dimers independently bind two plexin molecules and that signalling is critically dependent on the avidity of the resulting bivalent 2:2 complex (monomeric semaphorin binds plexin but fails to trigger signalling). In combination, our data favour a cell-cell signalling mechanism involving semaphorin-stabilized plexin dimerization, possibly followed by clustering, which is consistent with previous functional data. Furthermore, the shared generic architecture of the complexes, formed through conserved contacts of the amino-terminal seven-bladed β-propeller (sema) domains of both semaphorin and plexin, suggests that a common mode of interaction triggers all semaphorin–plexin based signalling, while distinct insertions within or between blades of the sema domains determine binding specificity.

Semaphorins are sub-divided into eight classes of cell-attached or secreted glycoproteins[4], characterized by an extracellular N-terminal sema domain followed by a cysteine rich PSI (plexin, semaphorin, integrin) domain, which form homodimers through substantial sema–sema domain interfaces[5,6]. Plexins are large type 1 single transmembrane-spanning cell surface receptors (Fig. 1a) and are divided into four classes[2,7]. Sequence analyses indicate an N-terminal sema domain, followed by a combination of three PSI domains and six IPT domains (Ig domain shared by plexins and transcription factors). For class B plexins the prototypic interaction of SEMA4D with PLXNB1 is implicated in migration and proliferation of neuronal, endothelial and tumour cells as well as in angiogenesis and axonal guidance[8,9]. Class 6 semaphorins (Sema6A, B, C and D) typically interact with class A plexins; Sema6A–PlxnA2 signalling controls axon guidance in the hippocampus and granule cell migration in the cerebellum[3,10,11]. A and B plexins have essentially identical cytoplasmic structures comprising a Ras GTPase-activating protein (GAP) topology with an inserted Rho GTPase-binding domain (RBD)[12,13]. Various mechanisms have been proposed for semaphorin-mediated activation of the plexin cytoplasmic region[6,14–17] but molecular level analyses of the effect of semaphorin-binding on plexin structure and oligomeric state are necessary to provide the paradigm for semaphorin–plexin signalling.

We have determined crystal structures of the phylogenetically distant[18] PLXNB1$_{1-2}$–SEMA4D$_{ecto}$ and PlxnA2$_{1-4}$–Sema6A$_{ecto}$ complexes

and the unliganded states of PlxnA2$_{1-4}$ and Sema6A$_{ecto}$ at 3.0, 2.2, 2.3 and 2.3 Å resolution, respectively (see Methods and Fig. 1). The PlxnA2$_{1-4}$ sema domain forms a seven-bladed β-propeller, elaborated with distinctive insertions more closely related to the Met receptor[19,20] than to the semaphorins[5,6] (see Supplementary Fig. 1), which is followed by a PSI domain (PSI-1), an IPT domain (IPT-1) and a second PSI domain (PSI-2) (Fig. 1b and Supplementary Fig. 2) that together form a stalk pointing away from the sema domain. The crystal packing provides no evidence for oligomerization and PlxnA2$_{1-4}$ is monomeric in solution to concentrations of at least 29 μM (Fig. 1c and Supplementary Fig. 3). Unbound SEMA4D$_{ecto}$[5] and Sema6A$_{ecto}$ form dimers in the crystal and in solution (Supplementary Figs 3 and 4). The PlxnA2$_{1-4}$–Sema6A$_{ecto}$ and PLXNB1$_{1-2}$–SEMA4D$_{ecto}$ complexes are structurally similar. In both, the semaphorins and plexins interact in a 'head-to-head' fashion through their sema domains such that the semaphorin dimer brings together two plexin monomers to form a symmetric 2:2 complex (Fig. 1d and Supplementary Figs 2 and 4). Each of the plexins almost exclusively interacts one-to-one with a separate semaphorin chain and the two plexins diverge from the semaphorin dimer without interacting with each other. This architecture positions semaphorin and plexin C termini in a *trans* (anti-parallel) arrangement suitable for signalling between apposing cell surfaces.

Neither the semaphorin ectodomains of SEMA4D$_{ecto}$ and Sema6A$_{ecto}$ nor the four domain N-terminal portion of PlxnA2$_{1-4}$ undergo large conformational changes upon complex formation (Supplementary Fig. 5). Superpositions of the bound and unbound sema domains of SEMA4D$_{ecto}$, Sema6A$_{ecto}$ and PlxnA2$_{1-4}$ reveal no significant structural differences (Cα atom root mean squared deviations (r.m.s.d.) of 0.85, 0.71 and 0.59 Å, respectively). Conformational changes at the semaphorin–plexin interfaces are very limited (Supplementary Fig. 5a), indicating that the binding surfaces are essentially preformed in solution. Small differences in the orientation of the two subunits of the semaphorin dimer are apparent on comparison of the unbound and bound crystal structures for both SEMA4D$_{ecto}$ and Sema6A$_{ecto}$; however, there is no evidence for a re-orientation characteristic of complex formation and the dimer interface remains predominantly unchanged (Supplementary Fig. 5b). A similar low level of orientational flexibility is observed in the linkage of the sema domains with sema-phorin PSI and PlxnA2$_{1-4}$ PSI-1 domains (Supplementary Fig. 5b, c), as well as between the plexin IPT and PSI domains (Supplementary Fig. 5c); however, PlxnA2$_{1-4}$ PSI-2 is disordered in the complex crystal. Thus the semaphorin ectodomains appear to be relatively rigid structures, but full-length plexins may be more flexible.

The representative semaphorin 4–plexin B and semaphorin 6–plexin A complexes have the same overall shape, each of the four protein chains is located in a comparable position and orientation and the interface is located at equivalent positions on the sema domains (Figs 1d and 2), although the plexin adopts a slightly different orientation in the two complexes (Supplementary Fig. 6a). The sema domains are highly conserved within the semaphorin and plexin families both in structure (PLXNB1$_{1-2}$ and PlxnA2$_{1-4}$ r.m.s.d. of 1.64 Å for 390

---

[1]Division of Structural Biology, Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford OX3 7BN, UK. [2]Smurfit Institute of Genetics and Institute of Neuroscience, Trinity College Dublin, Dublin 2, Ireland.
*These authors contributed equally to this work.
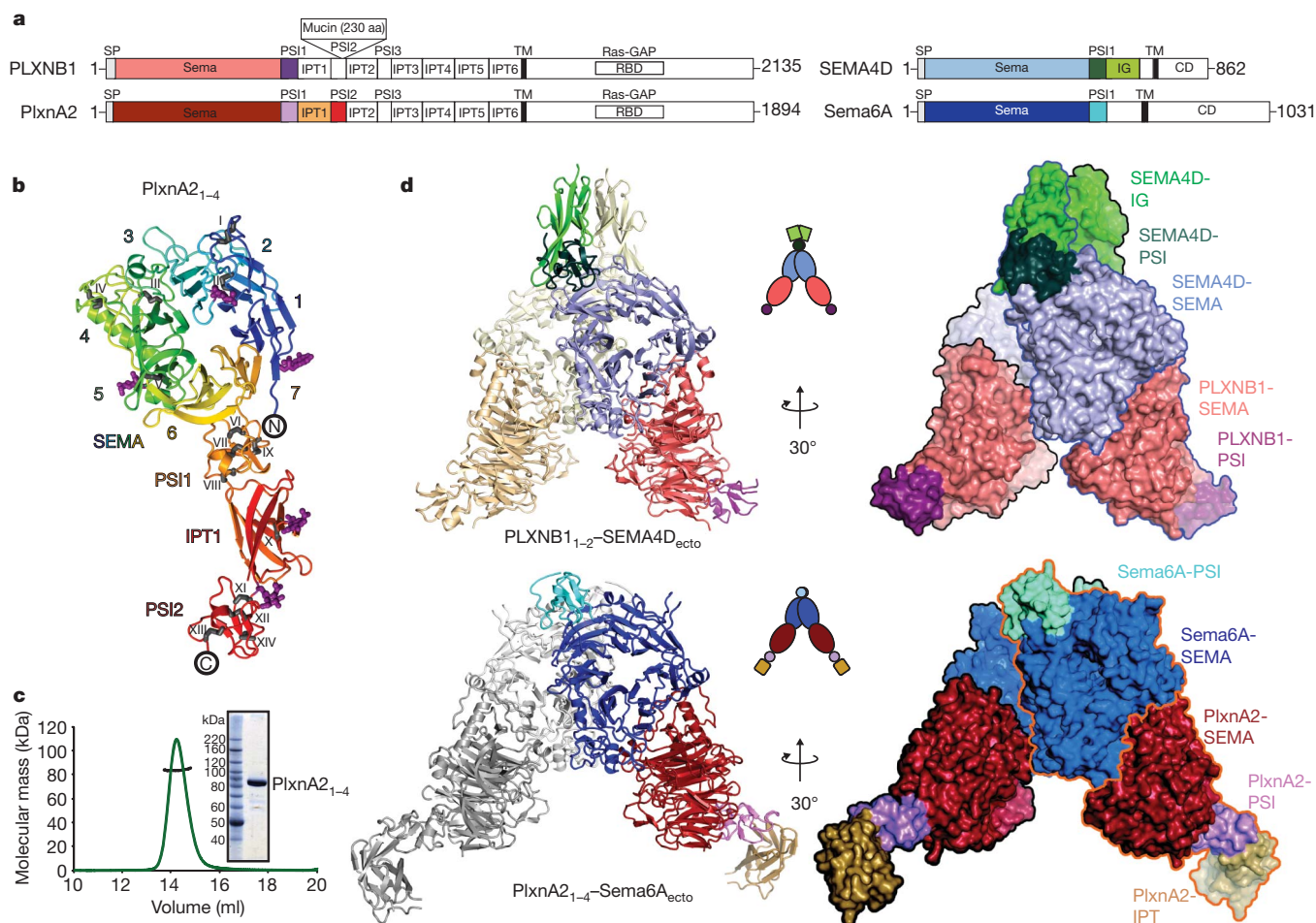
Figure 1 | The semaphorin–plexin complexes share a common architecture. a, Schematic domain organization of human PLXNB1, mouse PlxnA2, human SEMA4D and mouse Sema6A. PLXNB1 contains an additional mucin-like domain inserted into the PSI2 domain. SP, signal peptide; TM, transmembrane. The domains included in the crystallization constructs are coloured. **b,** Ribbon representation of PlxnA2$_{1-4}$ 'rainbow' colour ramped from blue (N terminus) to red (C terminus) with the β-propeller blades numbered. N-linked glycans are shown in magenta ball-and-stick representation and the 14 disulphide bridges (black stick presentation) are marked with Roman numbering. **c,** Multi-angle light scattering indicates an experimental molecular mass (black line) of 83.7 ± 0.8 kDa for PlxnA2$_{1-4}$ (green line; elution profile, axis not shown) as observed by SDS–PAGE (inset) and in agreement with the theoretical molecular mass for a monomer (85 kDa). **d,** Ribbon representation (left panel), cartoon drawing (middle panel) and surface representations with individual protein chains indicated by an outline (right panel) of the PLXNB1$_{1-2}$–SEMA4D$_{ecto}$ and PlxnA2$_{1-4}$–Sema6A$_{ecto}$ complexes. Domains are coloured as in Fig. 1a.

equivalent Cα atoms; SEMA4D$_{ecto}$ and Sema6A$_{ecto}$ r.m.s.d. of 1.73 Å r.m.s.d. for 427 Cα pairs; see also Supplementary Figs 1 and 6b) and in overall charge distribution (semaphorins positive and plexins negative; Fig. 2b). Furthermore, sequence alignments across the respective families show conserved surface residues cluster at the complex interface consistent with this mode of binding being common to all semaphorin–plexin interactions (Fig. 2b).

A single semaphorin and plexin chain together form an extensive, slightly discontinuous interface, burying 2,500 and 2,060 Å^2 in the PLXNB1$_{1-2}$–SEMA4D$_{ecto}$ and PlxnA2$_{1-4}$–Sema6A$_{ecto}$ complexes, respectively, and comprising a mixture of hydrophobic and hydrophilic (complementarily charged) patches (Fig. 2). Previous functional studies have implicated the semaphorin and plexin sema domains in complex formation and signalling[2,21]. The complex structures reveal that the same 'edge on' face of the β-propeller is used for the interaction by both families of molecules (Fig. 2a). This interaction site is predominantly formed by distinctive insertions in or between blades 1 to 5 of the semaphorin and plexin sema domains, with very few residues contributing directly from the blades of the standard β-propeller architecture (Supplementary Figs 2 and 4). The two most prominent insertions, a ~20 residue loop between blades 1 and 2 (β1D-β2A) and

a ~70 residue insert within blade 5 (first termed the extrusion in semaphorins[5]), are both present in semaphorins and plexins but with different lengths and conformations (Fig. 2a and Supplementary Figs 2 and 4). These novel structural features interact in an anti-parallel (*trans*) fashion (the semaphorin blade 1-2 loop with the plexin blade 5 insertion and vice versa) leading to a twofold arrangement of the interacting sema domains. Finer-grained differences in the interface-forming insertions (Supplementary Fig. 6c), particularly between the plexins PLXNB1$_{1-2}$ and PlxnA2$_{1-4}$, seem sufficient to discriminate against non-cognate complex formation.

The architecture of the complex, with a semaphorin dimer binding two plexin molecules indicates that bivalency is likely to have an important role in semaphorin–plexin interactions. Indeed a strong bivalency effect is observed in surface plasmon resonance (SPR) equilibrium experiments for both the PLXNB1$_{1-2}$–SEMA4D$_{ecto}$ and PlxnA2$_{1-4}$–Sema6A$_{ecto}$ interaction (Fig. 3a and Supplementary Fig. 7). Consistent with bivalency, when plexin is coupled to the Biacore chip, the semaphorin–plexin interaction does not follow simple 1:1 binding and the apparent affinity increases at higher plexin coupling densities due to the greater potential for bivalent interaction (Fig. 3a and Supplementary Fig. 7a, c)[22]. To facilitate direct comparison with previously reported
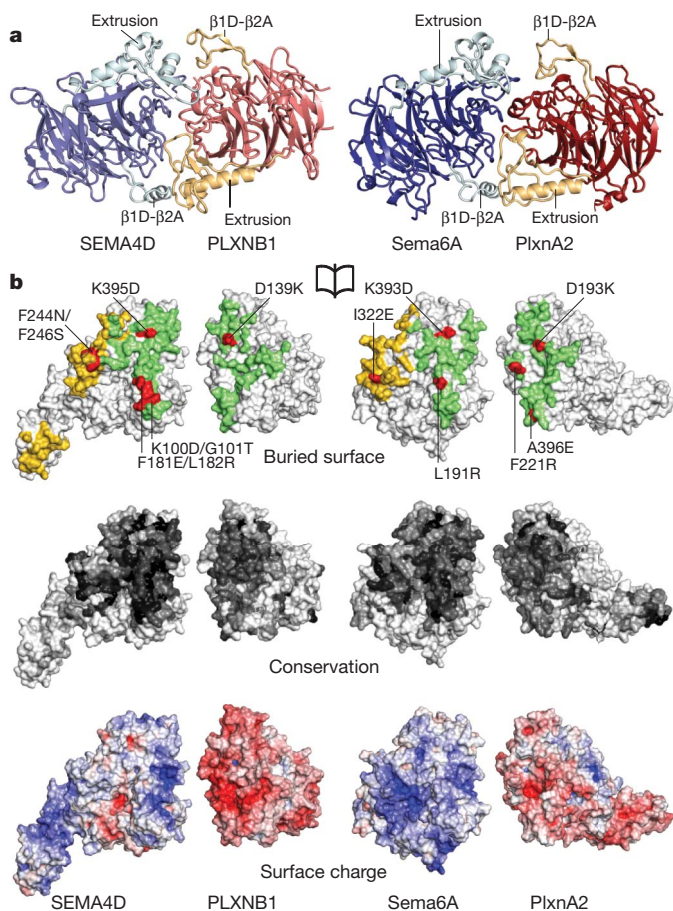
Figure 2 | Similar characteristics mediate the semaphorin–plexin interactions. a, Ribbon representation of the pseudo twofold arrangement of the interacting semaphorin and plexin sema domains. **b**, An opened view showing the semaphorin–plexin interface (green), the semaphorin homodimer interface (yellow) and interface mutants used in biophysical and cellular assays (red) (top panel). Semaphorin and plexin are colour-coded according to residue conservation (from non-conserved, white, to conserved, black) based on alignments containing sequences from all vertebrate semaphorin and plexin classes (middle panel). Semaphorin and plexin coloured by electrostatic potential from red ($-8\,k_bT/e_c$) to blue ($8\,k_bT/e_c$) (bottom panel). In both complexes the interface consists of conserved complementary charged patches.

assays of semaphorin–plexin binding affinities the SPR experiments were repeated with Fc-tagged dimerized Sema6A$_{ecto}$ (Supplementary Fig. 7g) and, although it is difficult to reach an exclusively bivalent interaction in SPR experiments[22], apparent affinities of up to 15 nM were measured, close to the nM values observed for cell-based assays with Fc-tagged semaphorins[2,3]. As expected, the reversed interaction (soluble plexin binding to chip coupled semaphorin; only performed for PlxnA2$_{1–4}$–Sema6A$_{ecto}$) is monovalent (1:1 binding), with a greater than 40-fold decrease in binding affinity (K_d = 2.3 μM), and independent of the density of semaphorin on the chip (Fig. 3a and Supplementary Fig. 7d). In order to dissect further the role of dimeric semaphorin in the bivalent interaction the homodimer interface was disrupted by mutagenesis to produce monomerized semaphorin (Fig. 3b and Supplementary Fig. 8). The bivalent interaction observed for plexin-coupled chips is converted to a monovalent interaction; the monomerized semaphorins giving K_d values of 5.5 μM for SEMA4D$_{ecto}$(F244N/F246S) and 1.3 μM for Sema6A$_{ecto}$(I322E) (Supplementary Fig. 7b, e) which is similar to the reversed 1:1 interaction observed for both dimeric and monomerized semaphorin-coupled chips (Supplementary Fig. 7d, f). Earlier studies have shown that semaphorin dimers are necessary for activity[21,23,24], possibly due to avidity. However, the monomerized SEMA4D$_{ecto}$(F244N/F246S) is not capable of triggering the collapse response in the well
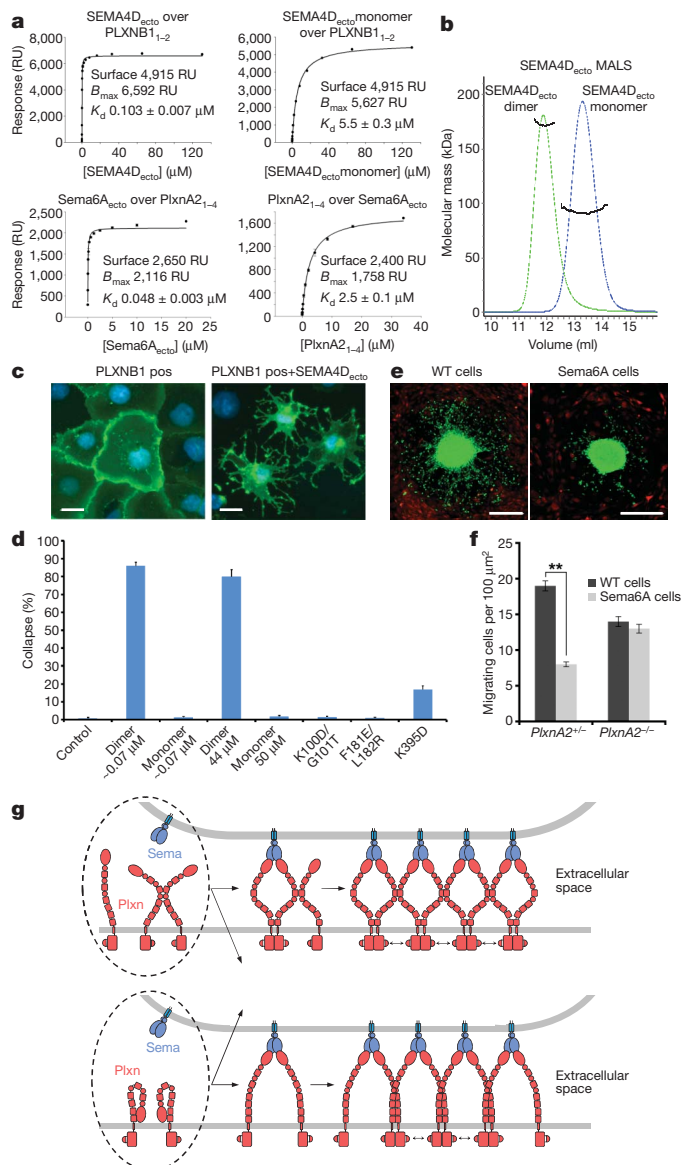
Figure 3 | Bivalent interaction is critical for semaphorin–plexin-induced cell-cell signalling. a, SPR equilibrium experiments of PLXNB1$_{1–2}$–SEMA4D$_{ecto}$ (top panels; wild-type SEMA4D$_{ecto}$ and monomerized SEMA4D$_{ecto}$(F244N/F246S)) and PlxnA2$_{1–4}$–Sema6A$_{ecto}$ (bottom panels; Sema6A$_{ecto}$ over PlxnA2$_{1–4}$ and reversed orientation, see also Supplementary Fig. 7). RU, response units. **b**, MALS analyses indicate molecular masses (black lines) of 174 ± 2 kDa and 92 ± 1 kDa for SEMA4D$_{ecto}$ and SEMA4D$_{ecto}$(F244N/F246S), respectively (elution profiles; green and blue lines, axis not shown). **c, d**, Cos-7 cell collapse assay showing representative images of non-collapsed cells (**c**, left panel) and SEMA4D$_{ecto}$-induced collapsed cells (**c**, right panel). Scale bar, 40 μm. **e**, EGL explants (green) grown on NIH3T3 cells (red) without (left panel) or with (right panel) Sema6A expression show the migration of post-mitotic granular neurons. WT, wild type. Scale bar, 200 μm. **f**, Quantification of migrating post-mitotic EGL explants of either *PlxnA2$^{+/-}$* or *PlxnA2$^{-/-}$* mice grown on wild type (shaded) or Sema6A expressing cells (hatched). (**P ≤ 0.005 by unpaired t-test). **g**, Model for semaphorin-stabilized plexin signalling. Binding of semaphorin stabilizes plexin dimerization, sufficient plexin ectodomain flexibility may enable plexin-to-plexin *cis* interaction in their membrane-proximal regions (upper panel) and seed further oligomerization. Possibly, dimerization is preceded by a 'switch-blade' conformational change in the plexin ectodomain (lower panel) exposing *cis* interaction sites leading to extracellular clustering. Two types of initial binding events (dotted enclosures) could result in the dimer and cluster architecture of either the upper or lower panel. The precise arrangement of the cytoplasmic region in the active state triggered by extracellular clustering cannot be specified.

established COS cell-based assay[25] even at a concentration ninefold above its K_d for PLXNB1 (Fig. 3c, d), showing that 1:1 binding of semaphorin to plexin is not enough to trigger signalling. Thus, semaphorin dimers facilitate bivalent interaction, necessary for sufficiently tight binding to plexins and crucial for semaphorin–plexin-induced cell-cell signalling.

Analyses of several semaphorin–plexin interface mutants (Fig. 2b and Supplementary Figs 2 and 4) in SPR and cell collapse assays are consistent with the crystallographically determined complex structures (Fig. 3d and Supplementary Fig. 9). Sema6A$_{ecto}$ mutant L191R binds PlxnA2$_{1-4}$ 5.5-fold weaker (Supplementary Fig. 9h). SEMA4D$_{ecto}$ mutants K100D/G101T and F181E/L182R and PlxnA2$_{1-4}$ mutants F221R and A396E all completely abolish binding (Supplementary Fig. 9f, g), and the SEMA4D$_{ecto}$ mutants do not collapse COS-7 cells (Fig. 3d). A PlxnA2(A396E) mutant, expressed in COS-7 cells, has also been reported to no longer bind Sema6A[11]. We further corroborated the interface by charge-reversal at a conserved salt bridge in both complexes; PLXNB1$_{1-2}$(D139K)–SEMA4D$_{ecto}$(K395D) (Supplementary Fig. 9c-e) and PlxnA2$_{1-4}$(D193K)–Sema6A$_{ecto}$(K393D) (Supplementary Fig. 9i-k). The semaphorin and plexin charge mutants bind over 200-fold and 9-fold weaker to their wild-type counterparts respectively, whereas binding is restored when the charge mutants are combined. Cerebellar granule cell migration assays on 3T3 cells directly demonstrate that contact with cells expressing full-length Sema6A reduces neuronal migration and that PlxnA2 is required for these neurons to respond to Sema6A, as previously suggested from genetic data (Fig. 3e, f)[3,11]. Mice harbouring the *PlxnA2* A396E mutation have previously been shown to exhibit defects in granule cell migration which are similar to *PlxnA2*$^{-/-}$ null mutant and *Sema6A*$^{-/-}$ mice[11], providing evidence *in vivo* for the importance of the Sema6A–PlxnA2 interaction defined by our structural data. The extensive class 3 semaphorin–plexin A interactions are distinctive in requiring members of the neuropilin family as co-receptors[15]. Previous studies have implicated two regions of the Sema3A sequence in function, sema domain blade 3 (ref. 21) and a K108N mutant that abolishes signalling *in vivo* (but does not disrupt neuropilin-1 binding)[26]. Both features map to the semaphorin–plexin interface that we observe consistent with direct Sema3–PlxnA interactions. Overall, the effects of semaphorin–plexin interface mutants *in vitro* and *in vivo* support the structural data and show that this interface is likely common to all semaphorin–plexin interactions.

In combination our above studies indicate that dimerization of the N-terminal domains of the plexin extracellular segment, resulting from bivalent semaphorin binding, is prerequisite for plexin signalling. After submission of this study, a similar architecture, and consequent role for plexin dimerization in signalling, was reported based on crystal structures of complexes of Sema7A and of a viral mimic with a sema-PSI fragment of PlxnC1 (ref. 27). We have, in addition, carried out biophysical analyses of eight domain and full-length extracellular (ten domain) constructs for PlxnA2 (PlxnA2$_{1-8}$ and PlxnA2$_{1-10}$) and of the entire cytoplasmic region of plexin B1 (PLXNB1$_{cyto}$) (Supplementary Fig. 10). Both PlxnA2$_{1-8}$ and PlxnA2$_{1-10}$ show increasing evidence of intermolecular interactions at higher concentrations consistent with the plexin ectodomain having some propensity for weak *cis*-interactions through membrane proximal domains 5–10 before semaphorin binding. In isolation, PLXNB1$_{cyto}$ remains monomeric at concentrations of at least 360 μM (Supplementary Fig. 10c) as shown previously at lower concentrations[13]. It has been reported that plexin activation can be induced by concurrent binding of intracellular Rnd1 and cluster-inducing antibodies[16,28] and that semaphorin binding induces clustering of plexins[15,16], possibly preceded by a conformational change in the plexin[14]. Full-length transmembrane PLXNB1 has weak *cis*-interaction which is further enhanced by SEMA4D[16]. Furthermore covalently oligomerized extracellular segment-deleted PLXNB1 is active independently of SEMA4D[16]. Semaphorin-stabilized plexin dimerization may seed further oligomerization through plexin-to-plexin *cis* interactions involving the membrane-proximal IPT domains or through intracellular

regions[16,28] (Fig. 3g). We observe some inter-domain flexion for the PlxnA2$_{1-4}$ IPT-1 and PSI-2 domains (Supplementary Fig. 5c) and additional flexibility may exist in the complete plexin extracellular region, analogous to that observed for the homologous Met receptor[19,20,29], to enable plexins to undergo a conformational change[14]. This change could expose *cis* interaction sites for extracellular clustering, providing an extra level of regulation to tightly control the activity of semaphorin–plexin induced cell-cell signalling.

METHODS SUMMARY

Human PLXNB1$_{1-2}$, human SEMA4D$_{ecto}$, mouse PlxnA2$_{1-4}$, PlxnA2$_{1-8}$, PlxnA2$_{1-10}$ and mouse Sema6A$_{ecto}$ (residues 20–535, 22–677, 35–703, 35–1040, 35–1231 and 19–571 respectively) were expressed in mammalian cells (CHO or HEK293) and human PLXNB1$_{cyto}$ (residues 1511–2135) was expressed in *Sf9* cells. All proteins were purified by immobilized metal ion affinity chromatography and size-exclusion chromatography. Crystals of human PLXNB1$_{1-2}$–SEMA4D$_{ecto}$ complex, mouse PlxnA2$_{1-4}$–Sema6A$_{ecto}$ complex and of mouse Sema6A$_{ecto}$ and PlxnA2$_{1-4}$ diffracted to 3.0, 2.2, 2.3 and 2.3 Å resolution, respectively, and structures were solved by molecular replacement and refined to final R_{work}/R_{free} values of 20.3/24.5%, 19.0/23.0%, 18.5/21.9% and 20.3/25.4%, respectively. The oligomerization state of PLXNB1$_{cyto}$, SEMA4D$_{ecto}$, PlxnA2$_{1-4}$, PlxnA2$_{1-8}$, PlxnA2$_{1-10}$, Sema6A$_{ecto}$, and semaphorin monomerizing mutants was determined with multi-angle light scattering (MALS) and analytical ultracentrifugation (AUC). SPR equilibrium binding experiments were performed on wild-type and semaphorin–plexin interface and semaphorin monomerizing mutant proteins. In these experiments the native membrane topology was mimicked by coupling of proteins via a carboxy-terminal-linked biotin label to streptavidin that was covalently precoupled to the surface. The ability of SEMA4D$_{ecto}$ mutants to activate PLXNB1 was tested in a cell collapse assay in COS-7 cells that expressed recombinant full length transmembrane human PLXNB1. Cerebellar granule cell migration assays on 3T3 cells were used to demonstrate directly that full-length Sema6A reduces neuronal migration and that PlxnA2 is required for neurons to respond to Sema6A.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

1. Kruger, R. P., Aurandt, J. & Guan, K. L. Semaphorins command cells to move. *Nature Rev. Mol. Cell Biol.* **6**, 789–800 (2005).
2. Tamagnone, L. *et al.* Plexins are a large family of receptors for transmembrane, secreted, and GPI-anchored semaphorins in vertebrates. *Cell* **99**, 71–80 (1999).
3. Suto, F. *et al.* Interactions between plexin-A2, plexin-A4, and semaphorin 6A control lamina-restricted projection of hippocampal mossy fibers. *Neuron* **53**, 535–547 (2007).
4. Kolodkin, A. L., Matthes, D. J. & Goodman, C. S. The semaphorin genes encode a family of transmembrane and secreted growth cone guidance molecules. *Cell* **75**, 1389–1399 (1993).
5. Love, C. A. *et al.* The ligand-binding face of the semaphorins revealed by the high-resolution crystal structure of SEMA4D. *Nature Struct. Biol.* **10**, 843–848 (2003).
6. Antipenko, A. *et al.* Structure of the semaphorin-3A receptor binding module. *Neuron* **39**, 589–598 (2003).
7. Winberg, M. L. *et al.* Plexin A is a neuronal semaphorin receptor that controls axon guidance. *Cell* **95**, 903–916 (1998).
8. Capparuccia, L. & Tamagnone, L. Semaphorin signaling in cancer cells and in cells of the tumor microenvironment–two sides of a coin. *J. Cell Sci.* **122**, 1723–1736 (2009).
9. Korostylev, A. *et al.* A functional role for semaphorin 4D/plexin B1 interactions in epithelial branching morphogenesis during organogenesis. *Development* **135**, 3333–3343 (2008).
10. Kerjan, G. *et al.* The transmembrane semaphorin Sema6A controls cerebellar granule cell migration. *Nature Neurosci.* **8**, 1516–1524 (2005).
11. Renaud, J. *et al.* Plexin-A2 and its ligand, Sema6A, control nucleus-centrosome coupling in migrating granule cells. *Nature Neurosci.* **11**, 440–449 (2008).
12. He, H., Yang, T., Terman, J. R. & Zhang, X. Crystal structure of the plexin A3 intracellular region reveals an autoinhibited conformation through active site sequestration. *Proc. Natl Acad. Sci. USA* **106**, 15610–15615 (2009).
13. Tong, Y. *et al.* Structure and function of the intracellular region of the plexin-B1 transmembrane receptor. *J. Biol. Chem.* **284**, 35962–35972 (2009).
14. Takahashi, T. & Strittmatter, S. M. Plexina1 autoinhibition by the plexin sema domain. *Neuron* **29**, 429–439 (2001).
15. Takahashi, T. *et al.* Plexin-neuropilin-1 complexes form functional semaphorin-3A receptors. *Cell* **99**, 59–69 (1999).
16. Oinuma, I., Katoh, H. & Negishi, M. Molecular dissection of the semaphorin 4D receptor plexin-B1-stimulated R-Ras GTPase-activating protein activity and neurite remodeling in hippocampal neurons. *J. Neurosci.* **24**, 11473–11480 (2004).

17. Tong, Y. *et al.* Binding of Rac1, Rnd1, and RhoD to a novel Rho GTPase interaction motif destabilizes dimerization of the plexin-B1 effector domain. *J. Biol. Chem.* **282,** 37215–37224 (2007).
18. Gherardi, E., Love, C. A., Esnouf, R. M. & Jones, E. Y. The sema domain. *Curr. Opin. Struct. Biol.* **14,** 669–678 (2004).
19. Stamos, J., Lazarus, R. A., Yao, X., Kirchhofer, D. & Wiesmann, C. Crystal structure of the HGF β-chain in complex with the Sema domain of the Met receptor. *EMBO J.* **23,** 2325–2335 (2004).
20. Niemann, H. H. *et al.* Structure of the human receptor tyrosine kinase met in complex with the *Listeria* invasion protein InlB. *Cell* **130,** 235–246 (2007).
21. Koppel, A. M., Feiner, L., Kobayashi, H. & Raper, J. A. A 70 amino acid region within the semaphorin domain activates specific cellular response of semaphorin family members. *Neuron* **19,** 531–537 (1997).
22. Müller, K. M., Arndt, K. M. & Pluckthun, A. Model and simulation of multivalent binding to fixed ligands. *Anal. Biochem.* **261,** 149–158 (1998).
23. Klostermann, A., Lohrum, M., Adams, R. H. & Puschel, A. W. The chemorepulsive activity of the axonal guidance signal semaphorin D requires dimerization. *J. Biol. Chem.* **273,** 7326–7331 (1998).
24. Koppel, A. M. & Raper, J. A. Collapsin-1 covalently dimerizes, and dimerization is necessary for collapsing activity. *J. Biol. Chem.* **273,** 15708–15713 (1998).
25. Turner, L. J. & Hall, A. Plexin-induced collapse assay in COS cells. *Methods Enzymol.* **406,** 665–676 (2006).
26. Merte, J. *et al.* A forward genetic screen in mice identifies Sema3A(K108N), which binds to neuropilin-1 but cannot signal. *J. Neurosci.* **30,** 5767–5775 (2010).
27. Liu, H. *et al.* Structural basis of Semaphorin-Plexin recognition and viral mimicry from Sema7A and A39R complexes with PlexinC1. *Cell* **142,** 749–761 (2010).
28. Toyofuku, T. *et al.* FARP2 triggers signals for Sema3A-mediated axonal repulsion. *Nature Neurosci.* **8,** 1712–1719 (2005).
29. Gherardi, E. *et al.* Structural basis of hepatocyte growth factor/scatter factor and MET signalling. *Proc. Natl Acad. Sci. USA* **103,** 4046–4051 (2006).

Author Contributions All authors contributed to the design of the project, data analysis and preparation of the manuscript. R.A.R. cloned, purified and performed SPR and MALS experiments on SEMA4D and PLXNB1 and crystallized and solved its complex structure. C.S. helped with the PLXNB1–SEMA4D complex structure solution. B.J.C.J. cloned, purified and performed SPR, AUC and MALS experiments on Sema6A and PlxnA2 and crystallized and solved the individual and complex structures. C.H.B. did the collapse assays, purified and performed AUC experiments on PLXNB1$_{cyto}$ and helped with other AUC experiments. F.P.-B. performed the granule cell migration assays.

Author Information Coordinates and structure factors for PLXNB1$_{1-2}$–SEMA4D$_{ecto}$, PlxnA2$_{1-4}$–Sema6A$_{ecto}$, Sema6A$_{ecto}$ and PlxnA2$_{1-4}$ have been deposited in the Protein Data Bank with succession numbers 3OL2, 3OKY, 3OKW and 3OKT, respectively. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to E.Y.J. (yvonne@strubi.ox.ac.uk).

METHODS

Production of SEMA4D$_{ecto}$, PLXNB1$_{1-2}$, Sema6A$_{ecto}$, PlxnA2$_{1-4}$, PlxnA2$_{1-8}$, PlxnA2$_{1-10}$ and PLXNB1$_{cyto}$. Human PLXNB1$_{1-2}$, human SEMA4D$_{ecto}$, mouse PlxnA2$_{1-4}$, PlxnA2$_{1-8}$, PlxnA2$_{1-10}$ and mouse Sema6A$_{ecto}$ (residues 20–535, 22–677, 35–703, 35–1040, 35–1231 and 19–571, respectively) were cloned into the pHLsec vector[30] in-frame with a C-terminal His-tag. For crystallization experiments SEMA4D$_{ecto}$ was produced in CHO lecR cells as described previously[5], PLXNB1$_{1-2}$ was expressed in HEK-293T cells[30] in the presence of kifunensine[31] and Sema6A$_{ecto}$ and PlxnA2$_{1-4}$ were expressed in HEK-293S GnTI⁻ cells[32]. For all other experiments proteins were expressed in HEK-293T cells (without glycosylation inhibitors) with the exception of Sema6A$_{ecto}$ and PlxnA2$_{1-4}$ that were expressed in HEK-293S cells for MALS and AUC experiments. Proteins were purified from buffer-exchanged medium by immobilized metal-affinity and size-exclusion chromatography. SEMA4D$_{ecto}$, Sema6A$_{ecto}$ and PlxnA2$_{1-4}$ were purified individually, whereas after metal-affinity purification PLXNB1$_{1-2}$ was mixed with purified SEMA4D$_{ecto}$ in a 2:1 ratio and the complex was subsequently purified by size-exclusion chromatography. The full-length intracellular domain of human PLXNB1 (residues 1511–2135; PLXNB1$_{cyto}$) was cloned into the pBac PAK9 vector (Clontech) in-frame with an N-terminal His-tag and used for transfection of 2 ml *Sf*9 cells (1×10^6 cells ml^{-1}). After five days the supernatant was collected and three rounds of virus amplification, each with a 1:100 dilution were carried out. For protein expression *Sf*9 cells with a density of 1.4×10^6 cells ml^{-1} were inoculated with the virus from the third round of amplification (1:10). After 3 days the cells were collected by centrifugation and resuspended in lysis buffer supplemented with protease inhibitors. Cells were lysed by sonication. PLXNB1$_{cyto}$ was purified from cleared lysate by immobilized metal-affinity and size-exclusion chromatography.

Crystallization and data collection. The PLXNB1$_{1-2}$–SEMA4D$_{ecto}$ complex was concentrated to 8.0 mg ml^{-1} in 10 mM Tris, pH 8.0 and 150 mM NaCl. Sema6A$_{ecto}$ and PlxnA2$_{1-4}$ were concentrated to 13.1 mg ml^{-1} and 6.8 mg ml^{-1}, respectively, both in 10 mM Hepes, pH 7.5 and 75 mM NaCl. Because we were unable to obtain well diffracting crystals of the glycosylated PlxnA2$_{1-4}$–Sema6A$_{ecto}$ complex both Sema6A$_{ecto}$ and PlxnA2$_{1-4}$ were deglycosylated with endoglycosidase F1 (ref. 31) and mixed at a molar ratio of 1:1 to a final concentration of 7.3 mg ml^{-1} before crystallization. For crystallization of individual proteins Sema6A$_{ecto}$ and PlxnA2$_{1-4}$ were not deglycosylated. Sitting drop vapour diffusion crystallization trials were set up using a Cartesian Technologies pipetting robot and consisted of 100 nl protein solution and 100 nl reservoir solution[33]. Crystallization plates were placed in a TAP Homebase storage vault maintained at 18 °C and imaged via a Veeco visualization system[34]. The PLXNB1$_{1-2}$–SEMA4D$_{ecto}$ complex crystallized in 0.1 M Tris, pH 7.0, 0.2 M calcium acetate, 6% v/v glycerol, 20% w/v PEG3000, the Sema6A$_{ecto}$–PlxnA2$_{1-4}$ complex in 0.12 M Mes, pH 6.0, 1% v/v ethyl acetate, 12.4% v/v 2-methyl-2,4-pentanediol, Sema6A$_{ecto}$ in 0.2 M di-ammonium hydrogen citrate, 6% w/v D-galactose, 20% w/v PEG3350, and PlxnA2$_{1-4}$ in 0.06 M HEPES, pH 7.0, 0.13 M magnesium chloride and 12.8% w/v PEG6000. Before diffraction data collection crystals were soaked in mother liquor supplemented with glycerol (25%, 15%, 20% and 20% v/v glycerol for PLXNB1$_{1-2}$–SEMA4D$_{ecto}$, PlxnA2$_{1-4}$–Sema6A$_{ecto}$, Sema6A$_{ecto}$ and PlxnA2$_{1-4}$, respectively) and subsequently flash-cooled in liquid nitrogen or in a cryo nitrogen gas stream. Data were collected at 100 K at Diamond beamline I03 (PLXNB1$_{1-2}$–SEMA4D$_{ecto}$, PlxnA2$_{1-4}$–Sema6A$_{ecto}$ and Sema6A$_{ecto}$) and at European Synchrotron Radiation Facility (ESRF) beamline ID23-1 (PlxnA2$_{1-4}$). Diffraction data were integrated and scaled with the HKL suite[35] (PLXNB1$_{1-2}$–SEMA4D$_{ecto}$) or with MOSFLM[36] and SCALA[37] in CCP4[38] (PlxnA2$_{1-4}$–Sema6A$_{ecto}$, Sema6A$_{ecto}$ and PlxnA2$_{1-4}$) (see Supplementary Table 1).

Structure determination and refinement. First we solved the structure of Sema6A$_{ecto}$ by molecular replacement in PHASER[39] using the structure of Sema3A[6] (Protein Database (PDB) code 1Q47) as a search model. This solution was subjected to one round of simulated annealing refinement in PHENIX[40] and subsequently re-built automatically by ARP/wARP[41] and completed by manual rebuilding in COOT[42] and refinement in PHENIX. The structure of the PlxnA2$_{1-4}$–Sema6A$_{ecto}$ complex was solved by molecular replacement in PHASER with the Sema6A$_{ecto}$ structure and the Met-receptor structure[20] (PDB code 2UZX) as search models, successively. This partial model was re-built automatically by ARP/wARP and BUCCANEER[43] and completed by several cycles of manual rebuilding in COOT and refinement in PHENIX. PlxnA2 domain PSI2 was omitted from the model due to disorder. Before manual rebuilding a mask was constructed around the putative PlxnA2$_{1-4}$ molecule in the PlxnA2$_{1-4}$–Sema6A$_{ecto}$ complex in CCP4, and the electron density inside the mask was used for molecular replacement in PHASER to solve the PlxnA2$_{1-4}$ structure. Using this solution an initial model was build automatically by ARP/wARP and completed by manual rebuilding in COOT and refinement in PHENIX. The structure of the PLXNB1$_{1-2}$–SEMA4D$_{ecto}$ complex was solved by molecular replacement in PHASER with the SEMA4D$_{ecto}$ structure[5] (PDB code 1OLZ) and the PlxnA2$_{1-4}$ structure as search models, successively.

This solution was re-built automatically by BUCCANEER and completed by several cycles of manual rebuilding in COOT and refinement in PHENIX and BUSTER[44]. Refinement statistics are given in Supplementary Table 1. All models were validated with MOLPROBITY[45]. Ramachandran statistics are as follows (favoured/disallowed (%)): PLXNB1$_{1-2}$–SEMA4D$_{ecto}$, 94.5/0.8, PlxnA2$_{1-4}$–Sema6A$_{ecto}$, 96.6/0.3, Sema6A$_{ecto}$ 96.3/0.1, PlxnA2$_{1-4}$ 98.0/0.3. Superpositions were calculated using SHP[46], electrostatics potentials were generated using APBS[47], alignments were calculated using ClustalW[48] and buried surface areas of protein–protein interactions were calculated using PISA[49]. Figures were produced using PyMOL (http://www.pymol.org/), ESPRIPT[50], ConSurf[51] and Adobe Photoshop (Adobe Systems) and Corel Draw (Corel Corporation).

Site-directed mutagenesis. Semaphorin–semaphorin dimer interface mutants SEMA4D$_{ecto}$(F244N/F246S) (introducing a glycosylation site) and Sema6A$_{ecto}$(I322E) and semaphorin–plexin interface mutants PLXNB1$_{1-2}$(D139K), PlxnA2$_{1-4}$(D193K), PlxnA2$_{1-4}$(F221R), PlxnA2$_{1-4}$(A396E), SEMA4D$_{ecto}$(K100D/G101T), SEMA4D$_{ecto}$ (F181E/L182R), SEMA4D$_{ecto}$(K395D), Sema6A$_{ecto}$(L191R) and Sema6A$_{ecto}$ (K393D) (see Fig. 2) were generated by a two-step overlapping PCR and cloned into the pHLsec mammalian expression vector, resulting in protein constructs with a C-terminal His$_6$ tag[30], a C-terminal BirA recognition sequence for biotinylation[52] or a C-terminal Fc tag for covalent dimerization. All mutant proteins were expressed in HEK-293T cells to ensure full glycosylation and were secreted at similar levels to the wild-type proteins. The stringent quality control mechanisms specific to the mammalian cell secretory pathway ensure that secreted proteins are correctly folded[53]. Mutant proteins were used for SPR, cell collapse, AUC and MALS experiments.

Multi-angle light scattering. MALS experiments were performed during size exclusion chromatography on either an analytical Superdex S200 10/30 column (GE Heathcare) or a TSK-Gel G3000SWXL column (Tosoh) with online static light-scattering (DAWN HELEOS II, Wyatt Technology), differential refractive index (Optilab rEX, Wyatt Technology) and Agilent 1200 UV (Agilent Techologies) detectors. Proteins had previously been purified by size-exclusion chromatography. Data were analysed using the ASTRA software package (Wyatt Technology).

Analytical ultracentrifugation. Sedimentation velocity experiments were performed using an Optima Xl-I analytical ultracentrifuge (Beckman). Purified SEMA4D$_{ecto}$, Sema6A$_{ecto}$ and PlxnA2$_{1-4}$ samples at different concentrations in 10 mM HEPES, pH 7.5 and 150 mM NaCl were centrifuged in double sector 12-mm centerpieces in a An-60 Ti rotor (Beckman) at 50,000 r.p.m. and 20 °C. Protein sedimentation was monitored by Rayleigh interference. Data were analysed using SEDFIT[54].

Surface plasmon resonance equilibrium binding studies. SPR equilibrium experiments were performed using a Biacore T100 machine (GE Healthcare) at 25 °C in 10 mM HEPES, pH 7.5, 150 mM NaCl, 0.005% (v/v) polysorbate 20. All proteins were homogeneous with full biological activity and underwent gel filtration in running buffer immediately before use. To mimic the native membrane insertion topology proteins were enzymatically biotinylated at an engineered C-terminal tag and attached via the biotin label to streptavidin that was covalently coupled to the surface[52]. To investigate the bivalent effect of semaphorin dimer binding to plexin three different concentrations of plexin were coupled to the surface[22]. The signal from experimental flow cells was corrected by subtraction of a blank and reference signal from a mock coupled flow cell in Scrubber2 (BioLogic). In all experiments analysed, the experimental trace returned to baseline after a regeneration step with 2 M MgCl$_2$ (Supplementary Figs 7 and 9). K_d and maximum analyte binding (B_{max}) values were obtained by nonlinear curve fitting of a 1:1 Langmuir interaction model (bound $= B_{max}/(K_d + C)$, where C is analyte concentration calculated as monomer) in SigmaPlot (Systat). In experiments with plexin coupled to the surface and dimeric semaphorins injected, binding did not fit well to a 1:1 model due to mixed bivalent and monovalent interaction. Apparent K_d values calculated in these experiments are therefore approximations. Nevertheless, apparent affinities of wild-type and mutant proteins determined at equal plexin coupling concentrations can be compared relative to each other.

Functional cell collapse assay. Cellular collapse assays were performed essentially as described[25]. Briefly, COS-7 cells were seeded on glass coverslips and transfected with human Plexin B1 carrying an N-terminal Flag-tag. Two days after transfection, cells were treated with medium containing secreted wild-type or mutant SEMA4D$_{ecto}$ and incubated for 30 min at 37 °C. Finally, the cells were fixed and stained with anti-Flag primary antibody (Sigma) and Alexa 488-labelled secondary antibody (Invitrogen). Cell nuclei were counterstained with DAPI (Invitrogen) and cells were visualized with a TE2000U fluorescence microscope (Nikon) equipped with an Orca CCD camera (Hamamatsu). Plexin B1-expressing cells were classified as collapsed or non-collapsed on the basis of reduced surface area. Each experiment was repeated twice and 2×200 cells were counted each time. Results are shown as mean with error bars representing standard error of the mean.

Creation of NIH3T3 stably transfected cell lines. NIH3T3 cells were transfected with pEF-1-dTomato alone or together with pCAGGS-flag-Semaphorin6A-full length. After three rounds of selection, the chosen clones for each line were selected based on two criteria: (1) the co-expression of both genes, (2) the proper targeting of Semaphorin6A to the plasma membrane. Both criteria were tested by fluorescent techniques using monoclonal antibody against the flag tag (clone M2, Sigma).

Explant cultures. C57/BL6 mice at postnatal day 5 (P5) were used from wild-type and Plexin-A2 knock-out (PA2 KO) mice models. Animals were sacrificed according to the Irish Department of Agriculture and European Ethical and Animal Welfare regulations. Cerebella were dissected out and sliced to get 200 µm cerebellum cortex slices. The external granular layer (EGL) was isolated from the cerebellar cortex and cut into 200–500 µm tissue pieces. EGL explants were cultured onto NIH3T3 monolayer cell lines stably transfected with different genes (dTomato, Flag–Semaphorin6A/dTomato). Co-cultures were grown in Dulbecco's modified Eagle's medium supplemented with L-glutamine, D-glucose, fetal bovine serum (FBS) and 3 M KCl, for 4 days in 5% CO_2, 95% humidity incubator at 37 °C.

Immunohistochemistry. Co-cultures were fixed in 4% PFA for 1 h. After several rinses with PBS, cultures were incubated with the monoclonal antibody against NeuN transcription factor (1:200, clone A60, Millipore) to mark the cell bodies of the post-mitotic cerebellar neurons. Secondary antibody conjugated to Alexa488 was used for further analysis in fluorescent microscopes.

Analysis and quantification. Immunostained co-cultures were examined on a Zeiss LSM-700 microscope. All the migrating post-mitotic cell bodies were counted as well as the explant area measured from each single co-culture picture using ImageJ image analysis software. Migration data were normalized as number of migrating cells for 100 µm^2 of explant area, and expressed as mean ± s.e.m. For each experimental condition 50–100 explants were used. Student's t-test was chosen for further statistical analysis.

30. Aricescu, A. R., Lu, W. & Jones, E. Y. A time- and cost-efficient system for high-level protein production in mammalian cells. *Acta Crystallogr. D* **62,** 1243–1250 (2006).
31. Chang, V. T. *et al.* Glycoprotein structural genomics: solving the glycosylation problem. *Structure* **15,** 267–273 (2007).
32. Reeves, P. J., Callewaert, N., Contreras, R. & Khorana, H. G. Structure and function in rhodopsin: high-level expression of rhodopsin with restricted and homogeneous N-glycosylation by a tetracycline-inducible N-acetylglucosaminyltransferase I-negative HEK293S stable mammalian cell line. *Proc. Natl Acad. Sci. USA* **99,** 13419–13424 (2002).
33. Walter, T. S. *et al.* A procedure for setting up high-throughput nanolitre crystallization experiments. Crystallization workflow for initial screening, automated storage, imaging and optimization. *Acta Crystallogr. D* **61,** 651–657 (2005).
34. Mayo, C. J. *et al.* Benefits of automated crystallization plate tracking, imaging, and analysis. *Structure* **13,** 175–182 (2005).
35. Otwinowski, Z. & Minor, W. Processing X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276,** 307–326 (1997).
36. Leslie, A. G. Recent changes to the MOSFLM package for processing film and image plate data. *Joint CCP4 + ESF-EAMCB Newsletter on Protein Crystallography* **26** (1992).
37. Evans, P. Scaling and assessment of data quality. *Acta Crystallogr. D* **62,** 72–82 (2006).
38. CCP4. The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. D* **50,** 760–763 (1994).
39. McCoy, A. J. *et al. Phaser* crystallographic software. *J. Appl. Cryst.* **40,** 658–674 (2007).
40. Adams, P. D. *et al. PHENIX:* building new software for automated crystallographic structure determination. *Acta Crystallogr. D* **58,** 1948–1954 (2002).
41. Perrakis, A., Morris, R. & Lamzin, V. S. Automated protein model building combined with iterative structure refinement. *Nature Struct. Biol.* **6,** 458–463 (1999).
42. Emsley, P. & Cowtan, K. *Coot:* model-building tools for molecular graphics. *Acta Crystallogr. D* **60,** 2126–2132 (2004).
43. Cowtan, K. The *Buccaneer* software for automated model building. 1. Tracing protein chains. *Acta Crystallogr. D* **62,** 1002–1011 (2006).
44. Blanc, E. *et al.* Refinement of severely incomplete structures with maximum likelihood in *BUSTER-TNT. Acta Crystallogr. D* **60,** 2210–2221 (2004).
45. Davis, I. W. *et al.* MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. *Nucleic Acids Res.* **35,** W375–W383 (2007).
46. Stuart, D. I., Levine, M., Muirhead, H. & Stammers, D. K. Crystal structure of cat muscle pyruvate kinase at a resolution of 2.6 Å. *J. Mol. Biol.* **134,** 109–142 (1979).
47. Baker, N. A., Sept, D., Joseph, S., Holst, M. J. & McCammon, J. A. Electrostatics of nanosystems: application to microtubules and the ribosome. *Proc. Natl Acad. Sci. USA* **98,** 10037–10041 (2001).
48. Larkin, M. A. *et al.* Clustal W and Clustal X version 2.0. *Bioinformatics* **23,** 2947–2948 (2007).
49. Krissinel, E. & Henrick, K. Inference of macromolecular assemblies from crystalline state. *J. Mol. Biol.* **372,** 774–797 (2007).
50. Gouet, P., Courcelle, E., Stuart, D. I. & Metoz, F. ESPript: analysis of multiple sequence alignments in PostScript. *Bioinformatics* **15,** 305–308 (1999).
51. Landau, M. *et al.* ConSurf 2005: the projection of evolutionary conservation scores of residues on protein structures. *Nucleic Acids Res.* **33,** W299–W302 (2005).
52. O'Callaghan C. A. *. et al.* BirA enzyme: production and application in the study of membrane receptor-ligand interactions by site-specific biotinylation. *Anal. Biochem.* **266,** 9–15 (1999).
53. Trombetta, E. S. & Parodi, A. J. Quality control and protein folding in the secretory pathway. *Annu. Rev. Cell Dev. Biol.* **19,** 649–676 (2003).
54. Schuck, P. Size-distribution analysis of macromolecules by sedimentation velocity ultracentrifugation and Lamm equation modeling. *Biophys. J.* **78,** 1606–1619 (2000).

LETTER

Structural basis for semaphorin signalling through the plexin receptor

Terukazu Nogi[1], Norihisa Yasui[1]†, Emiko Mihara[1], Yukiko Matsunaga[1], Masanori Noda[2], Naoya Yamashita[3], Toshihiko Toyofuku[4], Susumu Uchiyama[2], Yoshio Goshima[3], Atsushi Kumanogoh[4] & Junichi Takagi[1]

Semaphorins and their receptor plexins constitute a pleiotropic cell-signalling system that is used in a wide variety of biological processes, and both protein families have been implicated in numerous human diseases[1–4]. The binding of soluble or membrane-anchored semaphorins to the membrane-distal region of the plexin ecto-domain activates plexin's intrinsic GTPase-activating protein (GAP) at the cytoplasmic region, ultimately modulating cellular adhesion behaviour[5]. However, the structural mechanism underlying the receptor activation remains largely unknown. Here we report the crystal structures of the semaphorin 6A (Sema6A) receptor-binding fragment and the plexin A2 (PlxnA2) ligand-binding fragment in both their pre-signalling (that is, before binding) and signalling (after complex formation) states. Before binding, the Sema6A ectodomain was in the expected 'face-to-face' homodimer arrangement, similar to that adopted by Sema3A and Sema4D, whereas PlxnA2 was in an unexpected 'head-on' homodimer arrangement. In contrast, the structure of the Sema6A–PlxnA2 signalling complex revealed a 2:2 heterotetramer in which the two PlxnA2 monomers dissociated from one another and docked onto the top face of the Sema6A homodimer using the same interface as the head-on homodimer, indicating that plexins undergo 'partner exchange'. Cell-based activity measurements using mutant ligands/receptors confirmed that the Sema6A face-to-face dimer arrangement is physiologically relevant and is maintained throughout signalling events. Thus, homodimer-to-heterodimer transitions of cell-surface plexin that result in a specific orientation of its molecular axis relative to the membrane may constitute the structural mechanism by which the ligand-binding 'signal' is transmitted to the cytoplasmic region, inducing GAP domain rearrangements and activation.

Both semaphorins and plexins contain, at the amino terminus of their ectodomain, a ~500-residue sema domain followed by a short (~50 residues) plexin–semaphorin–integrin (PSI) domain. Those regions corresponding to the sema plus PSI segment of Sema6A (Sema6A$_{SP}$, residues 19–570) and PlxnA2 (PlxnA2$_{SP}$, residues 38–561) that mediate ligand–receptor interaction were first expressed in mammalian cell lines, and then purified and crystallized (Supplementary Fig. 1). Structures of Sema6A$_{SP}$ and PlxnA2$_{SP}$ were determined at 2.5 Å and 2.1 Å resolution, respectively (Fig. 1a, b, Supplementary Tables 1 and 2, and Supplementary Results). In both proteins, the sema domain displays a seven-bladed β-propeller fold very similar to previously determined structures of Sema3A[6], Sema4D[7] and Met[8]. In addition to the long 'extrusion' within blade 5 described previously[7] (hereafter called extrusion 2), we noted a second insertion between blades 1 and 2 that proved characteristic to all sema domains (hereafter called extrusion 1) (Fig. 1e).
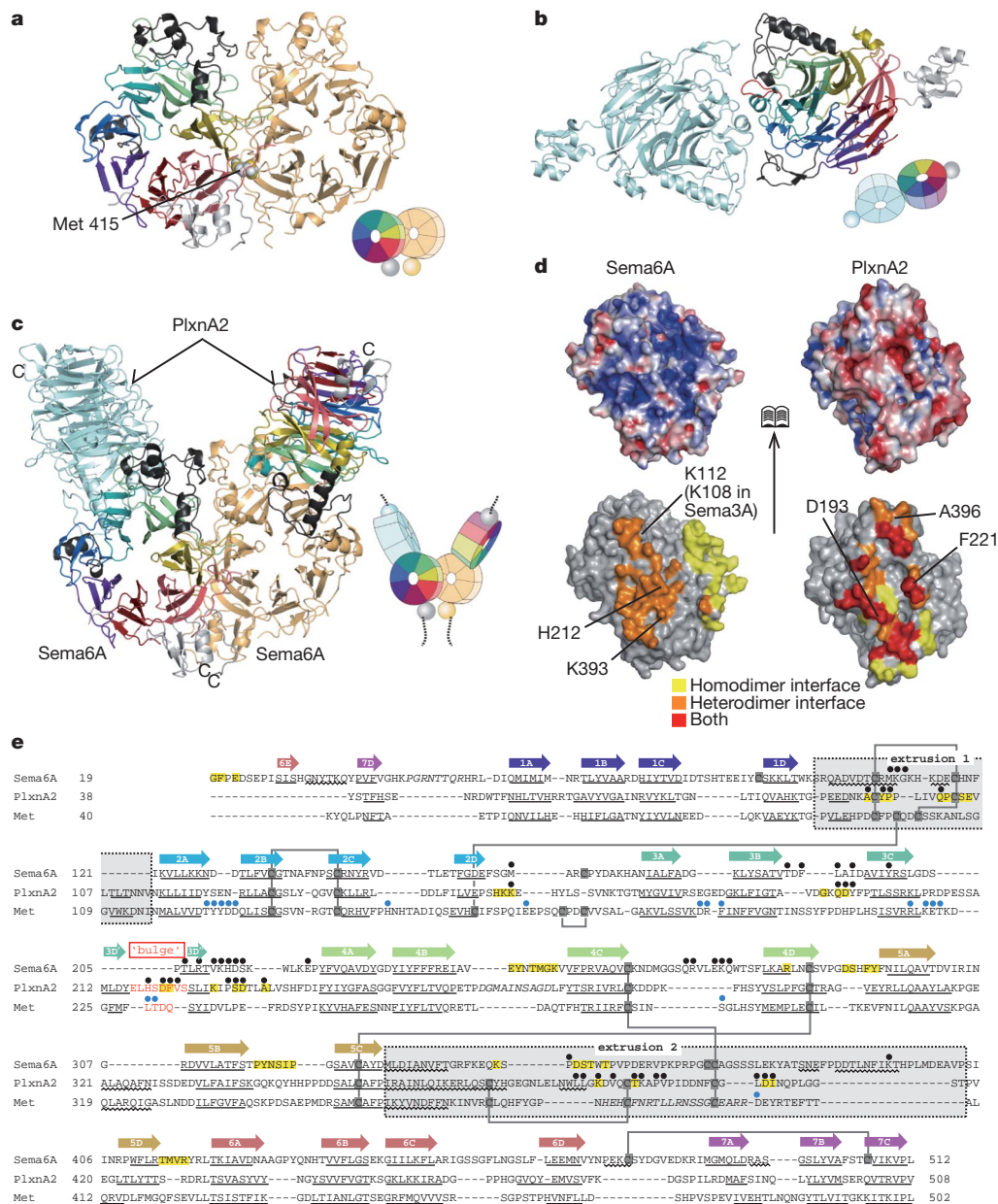
In the Sema6A$_{SP}$ crystal, monomers make contact with one another using the upper rim of the β-propeller, thereby assuming a 'face-to-face' dimer configuration (Fig. 1a). This dimeric configuration is essentially identical to that seen in the crystal structures of the Sema3A and Sema4D sema domains (Supplementary Fig. 2a). The location of the loops involved in the dimerization is precisely conserved among the three semaphorins, with the exception of the N-terminal region's participation in Sema6A (Supplementary Fig. 2b). Surprisingly, PlxnA2$_{SP}$ also assumes a dimeric configuration in the crystal, albeit with a markedly different mode compared to that observed in the semaphorin sema domains (Fig. 1b). The two PlxnA2$_{SP}$ fragments in the asymmetric unit are related by a non-crystallographic two-fold axis and interact with each other by using a flat surface located at the side of the β-propeller, exhibiting a 'head-on' configuration twisted orthogonally, in contrast to the face-to-face configuration observed in the known semaphorin structures. All the key residues involved in the dimerization are well conserved among the A-type plexin family (Supplementary Results and Supplementary Fig. 3), indicating the physiological relevance of the dimerization. Analytical ultracentrifugation sedimentation velocity experiments performed on the Sema6A$_{SP}$ protein confirmed that it does indeed form a dimer in solution with a dissociation constant (K_d) value of 3.5 μM (Supplementary Fig. 4). The dimerization affinity for PlxnA2$_{SP}$, however, was extremely low ($K_d > 300$ μM) and could not be definitively determined (Supplementary Fig. 5).

We next crystallized the Sema6A–PlxnA2 complex by mixing Sema6A$_{SP}$ and PlxnA2$_{SP}$ at an equimolar concentration, obtaining a structure at 3.6 Å resolution (Supplementary Results, Supplementary Table 3 and Supplementary Fig. 6). The Sema6A$_{SP}$ and PlxnA2$_{SP}$ molecules constitute a 2:2 complex in the crystal, which contained a crystallographic two-fold symmetry (Fig. 1c). The two Sema6A$_{SP}$ molecules in the complex formed the same face-to-face dimer as was observed in the plexin-free state (Supplementary Fig. 7a). On the other hand, the PlxnA2$_{SP}$ head-on homodimer was no longer present in the complex, and the two plexin molecules independently docked onto the two Sema6A monomers with their carboxy-terminal PSI domains emanating away diagonally. Despite their participation in different molecular interactions, there were no major changes in the structure of individual Sema6A$_{SP}$ and PlxnA2$_{SP}$ monomers, including the conformation of the loops at the interface, upon the complex formation (root mean squared deviation of 0.70 Å for Sema6A$_{SP}$ and 0.80 Å for PlxnA2$_{SP}$, respectively; Supplementary Figs 7 and 8). At the interface, the Sema6A side showed positively charged surface potentials whereas the PlxnA2 side was negatively charged, indicating that complex formation is driven mainly by electrostatic interactions (Fig. 1d).

We subsequently mutated a select number of interface residues on Sema6A to see whether these mutations disrupt plexin binding. The H212N mutation is expected to create a novel N212-D213-S214 glycosylation sequon and place a large carbohydrate obstacle at the heart of the interface (Fig. 1d and Supplementary Fig. 8). Another mutation, K393E, is expected to convert the electrostatic interaction between

Figure 1 | Crystal structure of Sema6A and PlxnA2 ectodomain fragments in pre-signalling and post-signalling states. a–c, Structures of the Sema6A face-to-face homodimer (**a**), PlxnA2 head-on dimer (**b**) and Sema6A–PlxnA2 2:2 complex (**c**). Individual propeller blades are coloured differently in one monomer. Arrangement of the toroidal propeller domains within the structure is schematically depicted in the cartoon next to each ribbon presentation. **d**, Open-book view of the Sema6A–PlxnA2 interface surface coloured by electrostatic potential (top panel) and by the residue-wise contribution to the interface (bottom panel). **e**, Structural alignment of the sema domains. Secondary structure elements are denoted by straight (strands) or wavy (helices) lines below each sequence. Residues in the homodimer interface are highlighted in yellow. Dots above the sequence indicate residues involved in the heteromeric interaction in the Sema6A–PlxnA2 complex (black) or in the Met–HGF complex (blue). Cysteines are shown with a grey background and disordered residues are shown in italics. Residues in a bulged insertion in the middle of strand 3D are shown in red.

Lys 393 and Asp 193 of the plexin into a repulsive one (Supplementary Fig. 8). When these mutations were introduced into the soluble Sema6A–alkaline phosphatase (AP) fusion protein, the binding of the mutant Sema6A–AP to the wild-type PlxnA2 expressed on HEK cells was markedly reduced or virtually absent (Fig. 2a), confirming the authenticity of the interface. As the same effect was observed when PlxnA4-expressing cells were used (data not shown), it is likely that the same binding site is used for the recognition of both plexin receptor subtypes.

Sema6A and Sema3A show similar biological activity on neurons and display the same plexin-type specificity[9,10], indicating that their plexin recognition sites are structurally conserved. In fact, Lys 112 in Sema6A that lies very close to the interface corresponds to the Sema3A

residue Lys 108, which has been shown to be critical in signalling events[11] (Fig. 1d and Supplementary Fig. 2b), further supporting this notion. We therefore mutated putative interface residues in Sema3A and tested their activity. As shown in Fig. 2b, Sema3A–AP proteins carrying interface mutations—including H216N (corresponding to Sema6A H212N glycosylation mutant) and R404E (corresponding to Sema6A K393E charge reversal mutant)—had lost their collapse-inducing activity, even though their binding to the high-affinity neuropilin-1 receptor remained intact (Fig. 2c). This indicates that the same interface is used by the neuropilin-bound Sema3A on the cell surface to bind to, and signal through, plexin receptors in neurons.

We next mutated key interface residues in the PlxnA2 sema domain and evaluated their ability to bind Sema6A. Ala 396 of PlxnA2 is
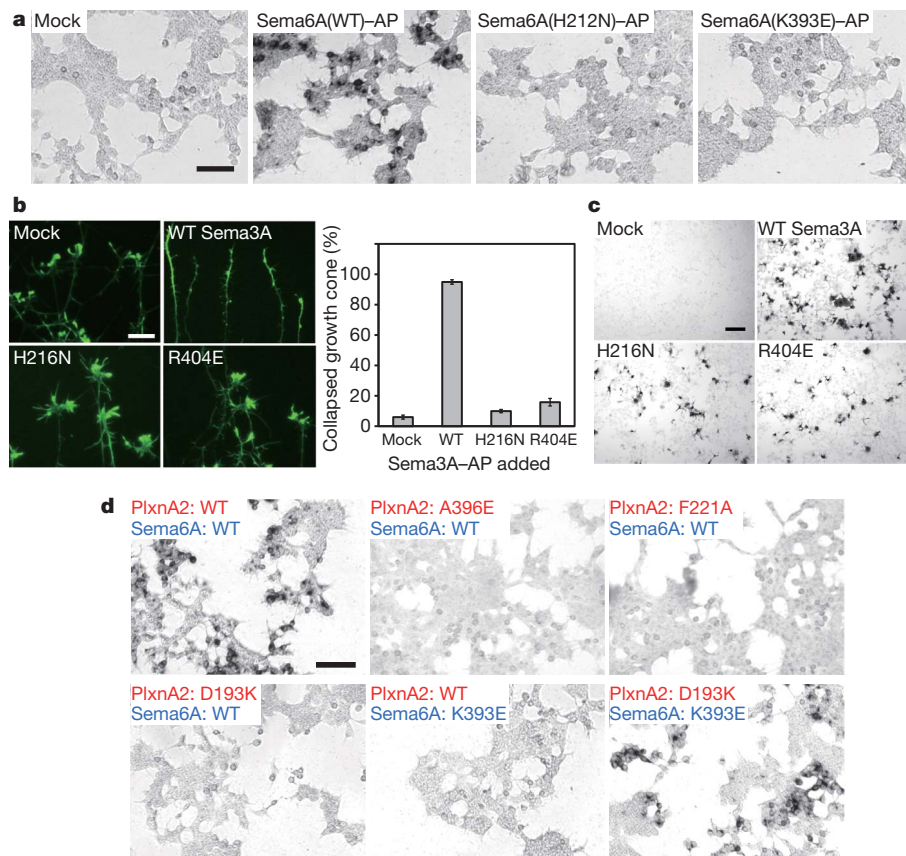
Figure 2 | Authenticity of the semaphorin–plexin interface is confirmed by mutational experiments. a, Binding of Sema6A–AP fusion protein to plexin. HEK293T cells transiently transfected with full-length wild-type PlxnA2 were incubated with either wild-type (WT) or mutant (H212N and K393E) Sema6A–AP at 100 nM for 90 min, fixed, and stained for AP activity. Scale bar, 100 μm. **b**, Activity loss of Sema3A mutants. Conditioned media containing wild-type, H216N or R404E Sema3A–AP were incubated with chick DRG neurons and visualized with Alexa488-conjugated phalloidin. Scale bar, 25 μm.

The percentage of collapsed growth cones was counted and expressed as the mean ± s.e.m. ($n = 9$). **c**, Neuropilin-1 binding activity of Sema3A mutants. COS-7 cells expressing mouse neuropilin-1 were incubated with the same set of conditioned medium used in **b** and the bindings were evaluated by AP activity staining. Scale bar, 200 μm. **d**, Effects of plexin mutations on Sema6A binding. HEK293T cells were transiently transfected with N-terminally Flag-tagged full-length PlxnA2 with indicated mutations and tested for the binding of either wild-type or the K393E mutant version of Sema6A–AP. Scale bar, 100 μm.

located at the heteromeric, but not the homodimeric, interface (Fig. 1d and Supplementary Fig. 8), in agreement with a finding that the A396E mutation in PlxnA2 causes a developmental abnormality in mice due to the lack of ligand binding[10]. We confirmed that the A396E mutant plexin expressed on HEK cells failed to support Sema6A–AP binding (Fig. 2d). Furthermore, an Ala substitution of Phe 221, which buries more than 170 Å^2 of the accessible surface in the complex, resulted in a negligible binding of Sema6A (Fig. 2d). Another mutation, D193K, which destroys salt bridge formation between Lys 393 of Sema6A, also eliminated the binding. This lack of Sema6A binding, however, did not stem from the impaired transportation of mutant plexins to the cell surface, as evidenced by the comparable level of anti-Flag immunostaining on the cell surface (Supplementary Fig. 9). Remarkably, when both charge-reversal mutants (that is, plexin D193K and Sema6A K393E) were combined, binding was completely recovered (Fig. 2d). These results strongly indicate that the current complex structure captures the genuine receptor–ligand interaction. Notably, PlxnA2 and Sema6A use an identical set of loops for heteromeric recognition (Fig. 1e, denoted by black dots). The resulting binding surface, in both proteins, is centred at blade 3, bounded by extrusions 1 and 2 at both sides, and located atop the propeller opposite the PSI domain. As Met also uses roughly the same region to interact with truncated HGF ligand (Fig. 1e, blue dots)[8], it is possible that the location of an important functional epitope is shared among all sema domains.

The strictly conserved dimeric configuration seen in all semaphorins prompted us to test the functional importance of the Sema6A dimer. Methionine 415 is located at the periphery of the dimerization interface,

pointing towards the corresponding residue in the partner molecule (Fig. 1a). Mutating Met 415 to Cys resulted in a covalent disulphide bond formation across the dimer interface in both cell-surface full-length Sema6A proteins (Fig. 3a), as well as in the soluble ectodomain fragment (Fig. 3b). We purified wild-type and dimeric (M415C) versions of the soluble ectodomain fragments of Sema6A and tested their ability to induce morphological changes in PlxnA4-expressing HEK cells. The locked dimer Sema6A (M415C) exhibited strong contraction activity towards those HEK cells stably expressing PlxnA4 over a concentration range of 10–300 nM (Fig. 3c), but not towards the control cells (data not shown). In contrast, very high concentrations (>3 μM) of the wild-type Sema6A ectodomain was required to elicit a modest level of cell contraction, indicating that it has >100-fold lower activity compared to the dimeric mutant. Dimerization-dependent collapse activity has already been reported using the Sema6A–Fc fusion protein[12]. Unlike the Fc or AP fusion strategy, which brings about dimerization of the recombinant proteins via Fc or AP moiety[13], our disulphide-bonded Sema6A dimerization did not permit dissociation and fixed the relative orientation of the two sema domains. Therefore, our results indicate that the heterotetrameric configuration of the Sema6A– PlxnA2 complex seen in the crystal structure represents a signalling-competent conformation maintained throughout the signal transduction process.

It has been shown that the cytoplasmic domain of plexin has the potential to form a homodimer[14]. It has also been postulated that at the extracellular region the sema domain can interact with the stalk region of the molecule[15]. Our results now reveal a third possible homophilic interaction mode in plexin: the sema–sema homodimerization.
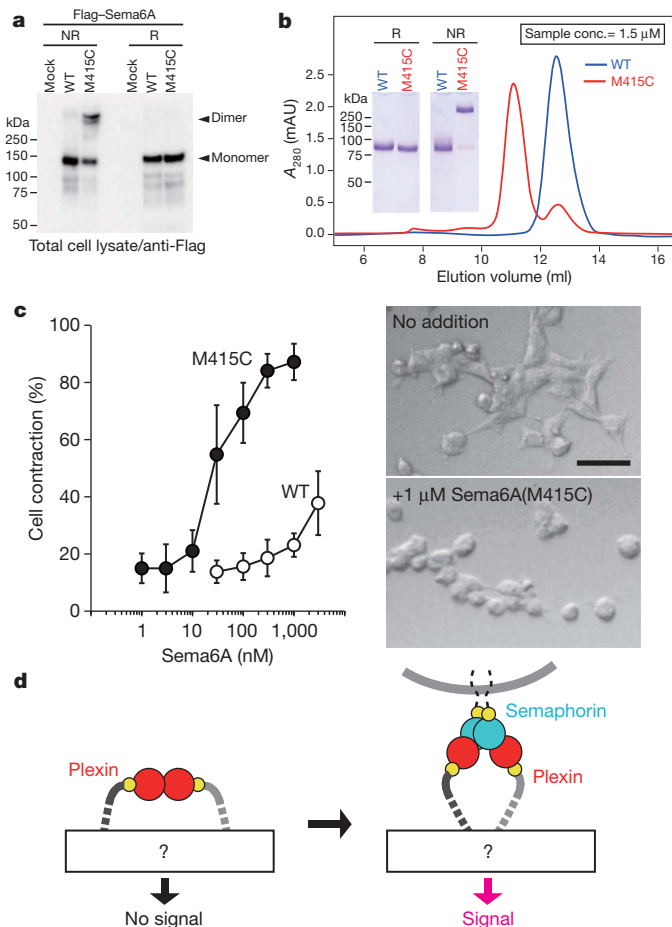
Figure 3 | The Sema6A face-to-face homodimer represents a signalling-competent active conformation. a, The dimer formation of Sema6A expressed on the cell surface. Lysates of HEK293T cells that had been transfected with empty vector (mock), Flag–Sema6A (WT), or Flag–Sema6A with the M415C mutation (M415C) were subjected to SDS–PAGE under reducing (R) or non-reducing (NR) conditions, followed by immunoblotting using anti-Flag antibody. **b**, Sema6A$_{SP}$ ectodomain fragments (WT or M415C) were subjected to gel filtration chromatography. The peak elution positions for the wild type and M415C mutant corresponded to 106 and 219 kDa, respectively. SDS–PAGE analysis (inset) confirmed the >90% formation of a disulphide-linked homodimer in the M415C mutant. **c**, Signalling activities of soluble Sema6A proteins. Purified Sema6A$_{SP}$ proteins (WT or M415C) were tested for their ability to induce contraction of HEK293T cells stably expressing PlxnA4. Representative images of cell morphologies both before and after the stimulation are also shown (right). Scale bar, 50 μm. **d**, Possible structural mechanism of semaphorin-induced plexin signalling. Transition from the head-on *cis* homodimer of plexin (left) to the semaphorin-engaged complex (right) changes the relative orientation of the plexin molecular axis. This conformational change is transmitted through the stalk region (thick dotted line) and alters the conformation (for example, dimerization state) of the cytoplasmic GAP domain, resulting in signal initiation. The closer positioning of the two plexin tails in the active conformation is drawn arbitrarily and should be taken as an example, because the association states of the transmembrane and cytoplasmic regions before and after receptor activation remain unknown.

Although affinity for the plexin sema domain homodimerization in solution was estimated to be extremely low (>300 μM), homomeric interactions with comparable degrees of affinity can mediate lateral receptor dimerization[16]. The head-on arrangement of the plexin *cis* dimer requires that the receptor axis be aligned in parallel to the membrane. Such a configuration would be possible in plexins, which have a long 'stalk region' consisting of six immunoglobulin domains interspersed by PSI domains found at the mobile domain boundaries both in integrins and in Met[17,18], thereby enabling the plexin to bend (or bow) its head. Similarly 'bowed' conformations have been postulated in Met as well[17].

Plexins on the resting cell surface assume an 'auto-inhibited' state, with their cytoplasmic GAP domain activity suppressed[19]. It is also accepted that ligand engagement at the extracellular side somehow activates GAP. Recently, crystal structures of the intracellular GAP domains of PlxnA3 and PlxnB1 have been reported by two groups[20,21]. Although it is still unclear how the activity of the GAP domain is structurally regulated (for example, by a monomer/dimer exchange or conformational changes within a single domain), our current structure clearly identifies the structural change that takes place at the extracellular side. In the resting state, plexin assumes an auto-inhibited conformation, possibly by structural constraints stemming from head-on dimerization (Fig. 3d, left). Upon semaphorin engagement, the orientation of the two plexin heads becomes more perpendicularly aligned to the membrane (Fig. 3d, right). This conformational change is then transmitted, through the long stalk and the transmembrane domain, to the cytoplasmic region, leading to activation of the GAP domain and/or recruitment of Rho family GTPases. Very recently, a structure determination of the Sema7A–PlxnC1 ectodomain complex was reported[22]. Although they did not solve the structure of the ligand-free PlxnC1, the complex structure was surprisingly similar to the Sema6A–PlxnA2 complex reported here, revealing a 2:2 stoichiometry with near-identical arrangements of each monomer. Although it remains possible that most of the cell-surface plexins in the resting state do not form a head-on dimer and the inactive phenotype is maintained by another type of mechanism[15], the structural conservation observed between the two semaphorin–plexin 'terminal' complexes is strongly indicative of the fundamental importance of this conformation in plexin signal transduction. More structural data are needed regarding the rest of the molecule, particularly the stalk region and the GAP domain, under different activation states, in order to understand fully the mechanism underlying semaphorin-induced plexin signal transduction. Such information may lead to the discovery of novel points of semaphorin signal intervention not limited to the receptor–ligand interface.

METHODS SUMMARY

The mouse Sema6A ectodomain fragment containing residues 19–570 was expressed as a human growth hormone (hGH) fusion protein in CHO lec 3.2.8.1 cells[23], whereas the mouse PlxnA2$_{SP}$ fragment (residues 38–561) containing the C-terminal TARGET tag was stably expressed using HEK293S GnT1⁻ cells as described previously[24]. Proteins were purified after removing the respective tag sequences and then crystallized. The Sema6A$_{SP}$ crystal with the highest diffraction quality was obtained in a buffer containing 22–24% (wt/vol.) polyethylene glycol (PEG) 1500 and 0.1 M Tris-Cl pH 7.0. PlxnA2$_{SP}$ was crystallized in a buffer containing 24–28% (wt/vol.) PEG 3350, 0–0.2 M NaCl and 0.1 M Tris-Cl pH 8.0–8.5. To determine the structure of the complex, Sema6A and PlxnA2 were mixed at an equimolar ratio and subjected to crystallization. This complex crystal grew in a solution containing 18–25% (wt/vol.) PEG 1000, 0.2–0.3 M MgCl$_2$, 0.1 M Na Cacodylate pH 5.5–6.5. The Sema6A structure was solved by a molecular replacement method using the coordinates of Sema4D (PDB 1OLZ). During molecular replacement phasing of the PlxnA2 crystal, the Met structure (PDB 1SHY) was used as a search model. The complex structure was determined by fitting the above-determined Sema6A and PlxnA2 structures. Single-isomorphous replacement with anomalous scattering (SIRAS) phasing with the Pt-derivative crystal was also incorporated during structural determination of the complex. Preparation of Sema6A–AP or Sema3A–AP fusion proteins, their cell binding analyses, and the growth cone collapse assay using explanted chick dorsal root ganglion (DRG) neurons were performed as previously described[25,26].

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

1. Tamagnone, L. & Comoglio, P. M. To move or not to move? Semaphorin signalling in cell migration. *EMBO Rep.* **5,** 356–361 (2004).
2. Zhou, Y., Gunput, R. A. & Pasterkamp, R. J. Semaphorin signaling: progress made and promises ahead. *Trends Biochem. Sci.* **33,** 161–170 (2008).
3. Suzuki, K., Kumanogoh, A. & Kikutani, H. Semaphorins and their receptors in immune cell interactions. *Nature Immunol.* **9,** 17–23 (2008).

4. Neufeld, G. & Kessler, O. The semaphorins: versatile regulators of tumour progression and tumour angiogenesis. *Nature Rev. Cancer* **8,** 632–645 (2008).
5. Pasterkamp, R. J. R-Ras fills another GAP in semaphorin signalling. *Trends Cell Biol.* **15,** 61–64 (2005).
6. Antipenko, A. *et al.* Structure of the semaphorin-3A receptor binding module. *Neuron* **39,** 589–598 (2003).
7. Love, C. A. *et al.* The ligand-binding face of the semaphorins revealed by the high-resolution crystal structure of SEMA4D. *Nature Struct. Biol.* **10,** 843–848 (2003).
8. Stamos, J., Lazarus, R. A., Yao, X., Kirchhofer, D. & Wiesmann, C. Crystal structure of the HGF β-chain in complex with the Sema domain of the Met receptor. *EMBO J.* **23,** 2325–2335 (2004).
9. Suto, F. *et al.* Plexin-A4 mediates axon-repulsive activities of both secreted and transmembrane semaphorins and plays roles in nerve fiber guidance. *J. Neurosci.* **25,** 3628–3637 (2005).
10. Renaud, J. *et al.* Plexin-A2 and its ligand, Sema6A, control nucleus-centrosome coupling in migrating granule cells. *Nature Neurosci.* **11,** 440–449 (2008).
11. Merte, J. *et al.* A forward genetic screen in mice identifies Sema3A(K108N), which binds to neuropilin-1 but cannot signal. *J. Neurosci.* **30,** 5767–5775 (2010).
12. Xu, X. M. *et al.* The transmembrane protein semaphorin 6A repels embryonic sympathetic axons. *J. Neurosci.* **20,** 2638–2648 (2000).
13. Le Du, M. H., Stigbrand, T., Taussig, M. J., Menez, A. & Stura, E. A. Crystal structure of alkaline phosphatase from human placenta at 1.8 Å resolution. Implication for a substrate specificity. *J. Biol. Chem.* **276,** 9158–9165 (2001).
14. Tong, Y. *et al.* Binding of Rac1, Rnd1, and RhoD to a novel Rho GTPase interaction motif destabilizes dimerization of the plexin-B1 effector domain. *J. Biol. Chem.* **282,** 37215–37224 (2007).
15. Takahashi, T. & Strittmatter, S. M. Plexina1 autoinhibition by the plexin sema domain. *Neuron* **29,** 429–439 (2001).
16. Chaudhry, C., Weston, M. C., Schuck, P., Rosenmund, C. & Mayer, M. L. Stability of ligand-binding domain dimer assembly controls kainate receptor desensitization. *EMBO J.* **28,** 1518–1530 (2009).
17. Niemann, H. H. *et al.* Structure of the human receptor tyrosine kinase met in complex with the *Listeria* invasion protein InlB. *Cell* **130,** 235–246 (2007).
18. Luo, B. H. & Springer, T. A. Integrin structures and conformational signaling. *Curr. Opin. Cell Biol.* **18,** 579–586 (2006).
19. Oinuma, I., Ishikawa, Y., Katoh, H. & Negishi, M. The Semaphorin 4D receptor Plexin-B1 is a GTPase activating protein for R-Ras. *Science* **305,** 862–865 (2004).
20. He, H., Yang, T., Terman, J. R. & Zhang, X. Crystal structure of the plexin A3 intracellular region reveals an autoinhibited conformation through active site sequestration. *Proc. Natl Acad. Sci. USA* **106,** 15610–15615 (2009).
21. Tong, Y. *et al.* Structure and function of the intracellular region of the plexin-B1 transmembrane receptor. *J. Biol. Chem.* **284,** 35962–35972 (2009).
22. Liu, H. *et al.* Structural basis of semaphorin-plexin recognition and viral mimicry from Sema7A and A39R complexes with plexinC1. *Cell* **142,** 749–761 (2010).
23. Stanley, P. Chinese hamster ovary cell mutants with multiple glycosylation defects for production of glycoproteins with minimal carbohydrate heterogeneity. *Mol. Cell. Biol.* **9,** 377–383 (1989).
24. Tabata, S. *et al.* A rapid screening method for cell lines producing singly-tagged recombinant proteins using the ''TARGET tag'' system. *J. Proteomics* **73,** 1777–1785 (2010).
25. Toyofuku, T. *et al.* Repulsive and attractive semaphorins cooperate to direct the navigation of cardiac neural crest cells. *Dev. Biol.* **321,** 251–262 (2008).
26. Goshima, Y. *et al.* A novel action of collapsin: collapsin-1 increases antero- and retrograde axoplasmic transport independently of growth cone collapse. *J. Neurobiol.* **33,** 316–328 (1997).

METHODS

Protein expression and purification. To express the Sema6A ectodomain fragment for crystallization, a DNA fragment corresponding to residues 19–570 was amplified from mouse Sema6A cDNA[25] and fused to the C terminus of the human growth hormone (hGH) minigene using a pSGHV0 vector[27]. The resultant plasmid was stably transfected into CHO lec 3.2.8.1 cells[23], and the clone with the highest secretion level was cultured in roller bottles (Corning). A hGH–Sema6A$_{SP}$ fragment was purified from the culture supernatants by Ni-NTA agarose chromatography and treated with tobacco etch virus protease to remove the hGH portion. The cleaved Sema6A$_{SP}$ fragment was further purified by gel filtration chromatography and concentrated to ~8 mg ml^{-1} before the crystallization trials. The mouse PlxnA2$_{SP}$ fragment containing the C-terminal TARGET tag was first stably expressed using HEK293S GnT1$^-$ cells and then affinity purified using P20.1-Sepharose as described previously[24].

Structure determination. Screening of crystallization conditions was performed by using mosquito (TTPLabtech). The Sema6A crystal was obtained from a solution containing 22–24% polyethylene glycol (PEG) 1500 and 0.1 M Tris-Cl pH 7.0. Cryoprotectant was prepared by mixing the reservoir solution and the ethylene glycol at a ratio of 4:1, resulting in a solution containing 20% ethylene glycol. Diffraction data were collected at Photon Factory BL-17A with an ADSC Quantum 270 CCD detector and at SPring-8 BL-41XU with a Rayonix MX225HE CCD detector, and then processed with HKL2000[28]. The crystal diffracted X-rays to 2.5 Å resolution and was found to belong to the space group $P2_1$ with unit cell dimensions of $a = 71.47$ Å, $b = 89.06$ Å, $c = 95.52$ Å and $\beta = 102.2°$. The initial phases were determined via the molecular replacement method with MOLREP[29], and the Sema4D structure (1OLZ) was used as a search model. Model building and refinement were performed with COOT[30] and REFMAC5[31] in which 5% of the reflections were excluded from the refinement. The crystallographic R-factor and the free R-factor were finally reduced to 21.1% and 27.8%, respectively, at 2.5 Å resolution. The quality of the final model was validated with MolProbity[32]. 95.08% of the amino acid residues were located in the favoured region of the Ramachandran plot and 0.29% were assigned as outliers.

The PlxnA2 crystal was obtained from a solution containing 24–28% PEG 3350, 0–0.2 M NaCl and 0.1 M Tris-Cl pH 8.0–8.5. Diffraction data were collected at Photon Factory BL-17A. Crystals diffracted X-rays to 2.1 Å resolution and were found to belong to the space group $P1$ with unit cell dimensions of $a = 55.70$ Å, $b = 60.68$ Å, $c = 95.39$ Å, $\alpha = 109.8°$, $\beta = 92.5°$ and $\gamma = 112.5°$. The structures of Sema3A (PDB 1Q47), Sema4D (PDB 1OLZ), Sema6A (PDB 3AFC) and the Met receptor (PDB 1SHY) were used in molecular replacement, with only the Met receptor yielding a clear solution. ARP/wARP[33] was used during the phase improvement and the crystallographic R-factor and the free R-factor were reduced to 20.5% and 25.8%, respectively, at 2.1 Å resolution. 97.56% of the residues were located in the favoured region of the Ramachandran plot and only Pro A-474 was assigned as an outlier.

The Sema6A–PlxnA2 complex crystal was obtained from a solution containing 18–25% PEG1000, 0.2–0.3 M MgCl$_2$ and 0.1 M Na Cacodylate pH 5.5–6.5, and was cryoprotected with a solution containing 25% PEG1000, 0.3 M MgCl$_2$, 0.1 M Na Cacodylate pH 6.0 and 20% ethylene glycol. Diffraction data were collected at Photon Factory BL-17A. The crystal diffracted X-rays up to 3.6 Å resolution and was found to belong to the space group $P6_122$ with unit cell dimensions of $a = b = 240.87$ Å, $c = 146.75$ Å. Molecular replacement phasing was performed using the Sema6A and PlxnA2 structures. The Pt-derivative crystal for SIRAS phasing was prepared by soaking the complex crystal in a solution containing 1 mM K$_2$PtCl$_4$, 23% PEG1000, 0.2 M MgCl$_2$, and 0.1 M Na Cacodylate pH 6.0 for 24 h. The Pt-derivative crystal diffracted X-rays to 4.5 Å resolution, and the data were collected at 1.07171 Å. The phases were calculated with SHLEXC/D[34] and SHARP/autoSHARP[35] and then improved with SOLOMON[36]. The crystallographic R-factor and free R-factor were finally reduced to 23.0% and 28.7%, respectively, at 3.6 Å resolution. 90.94% of the residues were located in the favoured region of the Ramachandran plot and 0.88% were assigned as outliers. Details of the data collection and refinement statistics for Sema6A, PlxnA2 and the complex are summarized in Supplementary Tables 1, 2 and 3, respectively. The

accessible surface area was calculated with AREAIMOL[37], and structural superposition was performed with SUPERPOSE[38]. Figures for the protein structures were prepared with PyMOL[39], in which electrostatic potentials were calculated with APBS[40]. For the structural details of the final models, see Supplemental Results.

Analytical ultracentrifugation. Measurements were performed with a ProteomeLab XL-I analytical ultracentrifuge (Beckman-Coulter) using An60 Ti rotor. For sedimentation velocity experiments on Sema6A$_{SP}$, runs were carried out at 42,000 r.p.m. at 20 °C using 12 mm aluminium double-sector cells loaded with various concentrations of Sema6A$_{SP}$ in buffer containing 20 mM HEPES, 150 mM NaCl, pH 7.5. For sedimentation equilibrium experiments on PlxnA2$_{SP}$, data were collected at 20 °C at 8, 16 and 48 μM in 5 mM Tris, 150 mM NaCl, pH 7.5, and at rotor speeds of 9,000, 12,000 and 15,000 r.p.m. Data were acquired with an ultraviolet absorbance detection system where appropriate wavelengths (246, 250, 258, 276, 280, 283 and 286 nm) were used depending on the concentration of the solution. Sedimentation velocity and sedimentation equilibrium data were analysed using the SEDFIT version 11.8 and SEDPHAT version 6.5 programs[41,42], respectively.

Biological assays. AP-ligand binding assays were performed as previously described[25]. To assess the biological activity of soluble Sema6A, a cell contraction assay was conducted. Briefly, HEK293T cells stably transfected with mouse PlxnA4 were seeded onto 10 mm × 10 mm glass coverslips coated with poly-L-lysine. After 24 h, the cells were incubated with varying concentrations of wild-type or M415C dimer mutant versions of Sema6A$_{SP}$ for 60 min at 37 °C. After fixing the samples with 4% paraformaldehyde in 10 mM PBS, pH 7.4 for 30 min at room temperature, the number of cells that underwent morphological changes from extended to round shapes was determined. The growth cone collapsing activity of the Sema3A–AP fusion proteins was assayed using explanted chick dorsal root ganglion (DRG) neurons as described previously[26].

27. Leahy, D. J., Dann, C. E. III, Longo, P., Perman, B. & Ramyar, K. X. A mammalian expression vector for expression and purification of secreted proteins for structural studies. *Protein Expr. Purif.* **20,** 500–506 (2000).
28. Otwinowski, Z. & Minor, W. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276,** 307–326 (1997).
29. Vagin, A. & Teplyakov, A. MOLREP: an automated program for molecular replacement. *J. Appl. Cryst.* **30,** 1022–1025 (1997).
30. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D* **60,** 2126–2132 (2004).
31. Murshudov, G. N., Vagin, A. A. & Dodson, E. J. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. D* **53,** 240–255 (1997).
32. Davis, I. W. *et al.* MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. *Nucleic Acids Res.* **35,** W375–W383 (2007).
33. Perrakis, A., Harkiolaki, M., Wilson, K. S. & Lamzin, V. S. ARP/wARP and molecular replacement. *Acta Crystallogr. D* **57,** 1445–1450 (2001).
34. Schneider, T. R. & Sheldrick, G. M. Substructure solution with SHELXD. *Acta Crystallogr. D* **58,** 1772–1779 (2002).
35. Bricogne, G., Vonrhein, C., Flensburg, C., Schiltz, M. & Paciorek, W. Generation, representation and flow of phase information in structure determination: recent developments in and around SHARP 2.0. *Acta Crystallogr. D* **59,** 2023–2030 (2003).
36. Abrahams, J. P. & Leslie, A. G. Methods used in the structure determination of bovine mitochondrial F1 ATPase. *Acta Crystallogr. D* **52,** 30–42 (1996).
37. Lee, B. & Richards, F. M. The interpretation of protein structures: estimation of static accessibility. *J. Mol. Biol.* **55,** 379–400 (1971).
38. Krissinel, E. & Henrick, K. Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. *Acta Crystallogr. D* **60,** 2256–2268 (2004).
39. DeLano, W. L. *The PyMOL Molecular Graphics System* (DeLano Scientific, 2002).
40. Baker, N. A., Sept, D., Joseph, S., Holst, M. J. & McCammon, J. A. Electrostatics of nanosystems: application to microtubules and the ribosome. *Proc. Natl Acad. Sci. USA* **98,** 10037–10041 (2001).
41. Schuck, P. Size-distribution analysis of macromolecules by sedimentation velocity ultracentrifugation and Lamm equation modeling. *Biophys. J.* **78,** 1606–1619 (2000).
42. Schuck, P. On the analysis of protein self-association by sedimentation velocity analytical ultracentrifugation. *Anal. Biochem.* **320,** 104–124 (2003).

LETTER

Maternal mRNA deadenylation and decay by the piRNA pathway in the early *Drosophila* embryo

Christel Rouget[1]*, Catherine Papin[1]*, Anthony Boureux[2], Anne-Cécile Meunier[1], Bénédicte Franco[1], Nicolas Robine[3], Eric C. Lai[3], Alain Pelisson[4] & Martine Simonelig[1]

Piwi-associated RNAs (piRNAs), a specific class of 24- to 30-nucleotide-long RNAs produced by the Piwi-type of Argonaute proteins, have a specific germline function in repressing transposable elements. This repression is thought to involve heterochromatin formation and transcriptional and post-transcriptional silencing[1–6]. The piRNA pathway has other essential functions in germline stem cell maintenance[7] and in maintaining germline DNA integrity[8–10]. Here we uncover an unexpected function of the piRNA pathway in the decay of maternal messenger RNAs and in translational repression in the early embryo. A subset of maternal mRNAs is degraded in the embryo at the maternal-to-zygotic transition. In *Drosophila*, maternal mRNA degradation depends on the RNA-binding protein Smaug and the deadenylase CCR4[11–13], as well as the zygotic expression of a microRNA cluster[14]. Using mRNA encoding the embryonic posterior morphogen Nanos (Nos) as a paradigm to study maternal mRNA decay, we found that CCR4-mediated deadenylation of *nos* depends on components of the piRNA pathway including piRNAs complementary to a specific region in the *nos* 3′ untranslated region. Reduced deadenylation when piRNA-induced regulation is impaired correlates with *nos* mRNA stabilization and translational derepression in the embryo, resulting in head development defects. Aubergine, one of the Argonaute proteins in the piRNA pathway, is present in a complex with Smaug, CCR4, *nos* mRNA and piRNAs that target the *nos* 3′ untranslated region, in the bulk of the embryo. We propose that piRNAs and their associated proteins act together with Smaug to recruit the CCR4 deadenylation complex to specific mRNAs, thus promoting their decay. Because the piRNAs involved in this regulation are produced from transposable elements, this identifies a direct developmental function for transposable elements in the regulation of gene expression.

In *Drosophila* embryos, Nos is expressed as a gradient that emanates from the posterior pole and organizes abdominal segmentation[15]. The majority of *nos* mRNA is distributed throughout the bulk cytoplasm, translationally repressed[16] and subsequently degraded during the first 2–3 h of development. This repression is essential for head and thorax segmentation[16,17]. A small amount of *nos* transcripts, localized at the posterior pole of the embryo, escapes degradation and is actively translated, giving rise to the Nos protein gradient. *nos* mRNA decay in the bulk cytoplasm depends on the CCR4–NOT deadenylation complex and its recruitment onto *nos* by Smaug (Smg). This contributes to translational repression in the bulk of the embryo and is required for embryonic antero-posterior patterning[13].

Smg has been suggested to be not the only activator of *nos* mRNA decay during early embryogenesis[11,12]. Zygotically expressed miRNAs have been reported to activate maternal mRNA deadenylation in zebrafish embryos[18] and decay in *Drosophila* embryos[14]. We investigated the potential involvement of other classes of small RNAs in mRNA deadenylation and decay before zygotic expression. Because piRNAs are expressed maternally in the germ line and are present in early embryos[19,20], we analysed the possible role of the piRNA pathway in maternal mRNA deadenylation. Piwi, Aubergine (Aub) and Ago3 are specific Argonaute proteins[1,3,21,22], Armitage (Armi) and Spindle-E (Spn-E) are RNA helicases, and Squash (Squ) is a nuclease[2,10,23,24] involved in piRNA biogenesis and function. Poly(A) test assays were performed to measure *nos* mRNA poly(A) tail length in embryos spanning 1-h intervals during the first 4 h of embryogenesis. In contrast to the progressive shortening of *nos* mRNA poly(A) tails observed in wild-type embryos correlating with mRNA decay during this period, *nos* poly(A) tail shortening was affected in embryos from females mutant for the piRNA pathway (herein referred to as mutant embryos) (Fig. 1a and Supplementary Figs 1a, 2 and 12). This defect in deadenylation correlated with higher amounts of *nos* mRNA in mutant embryos, as quantified by reverse transcription–quantitative PCR (RT–qPCR) (Fig. 1b). *In situ* hybridization revealed stabilized *nos* mRNA in the bulk cytoplasm of mutant embryos where it is normally degraded in the wild type (Fig. 1c and Supplementary Fig. 1b). Consistent with previous data showing that *nos* mRNA deadenylation is required for translational repression[13], defective deadenylation in mutant embryos resulted in the presence of ectopic Nos protein throughout the embryo (Fig. 1d and Supplementary Fig. 1c). The presence of Nos in the anterior region results in the repression of *bicoid* and *hunchback* mRNA translation and in affected head skeleton. Consistent with previously mentioned defects[7], we found that the *piwi[1]* mutant embryos that were able to produce a cuticle had head defects (Fig. 1e).

The piRNA pathway has a role during early oogenesis in preventing DNA damage, possibly through the repression of transposable element transposition. DNA double-strand breaks arising in mutants of the piRNA pathway correlate with affected embryonic axis specification, and this developmental defect is suppressed by mutations in the Chk2 DNA-damage signal transduction pathway[9,10]. We found that defects in *nos* mRNA deadenylation and decay observed in *aub* or *armi* mutants were not suppressed by Chk2 (*mnk[P6]*) mutations, indicating that these defects did not result from activation of the Chk2 pathway earlier during oogenesis (Supplementary Fig. 3a–c). Moreover, affected deadenylation of *nos* mRNA in piRNA pathway mutants did not depend on *oskar* (Supplementary Fig. 3d).

We addressed a potential direct role of the piRNA pathway in the regulation of *nos* mRNA deadenylation and decay in the embryo. Aub and Piwi accumulate in the pole plasm and in pole cells of the embryo[25,26]. However, we found lower levels of Aub and Piwi throughout the entire embryo (Fig. 2a and Supplementary Figs 4 and 5). Ago3 was also present throughout the embryo (Supplementary Fig. 6a, c). Aub and Ago3 were found in the cytoplasm and accumulated in discrete foci, a distribution similar to that of CCR4 and Smg (Fig. 2b and Supplementary Fig. 6b). CCR4 and Smg were reported to partially colocalize in small cytoplasmic foci[13]. Aub and Ago3 also partially

[1]mRNA Regulation and Development, Institute of Human Genetics, CNRS UPR1142, 141 rue de la Cardonille, Cedex 5, 34396 Montpellier, France. [2]CRBM, UMR5237, Université Montpellier II, CNRS, 1919 route de Mende, 34293 Montpellier, France. [3]Sloan-Kettering Institute, Department of Developmental Biology, 1017 Rockefeller Research Laboratories, 1275 York Avenue, New York, New York 10065, USA. [4]RNA Silencing and Control of Transposition, Institute of Human Genetics, CNRS UPR1142, 141 rue de la Cardonille, Cedex 5, 34396 Montpellier, France.
*These authors contributed equally to this work.
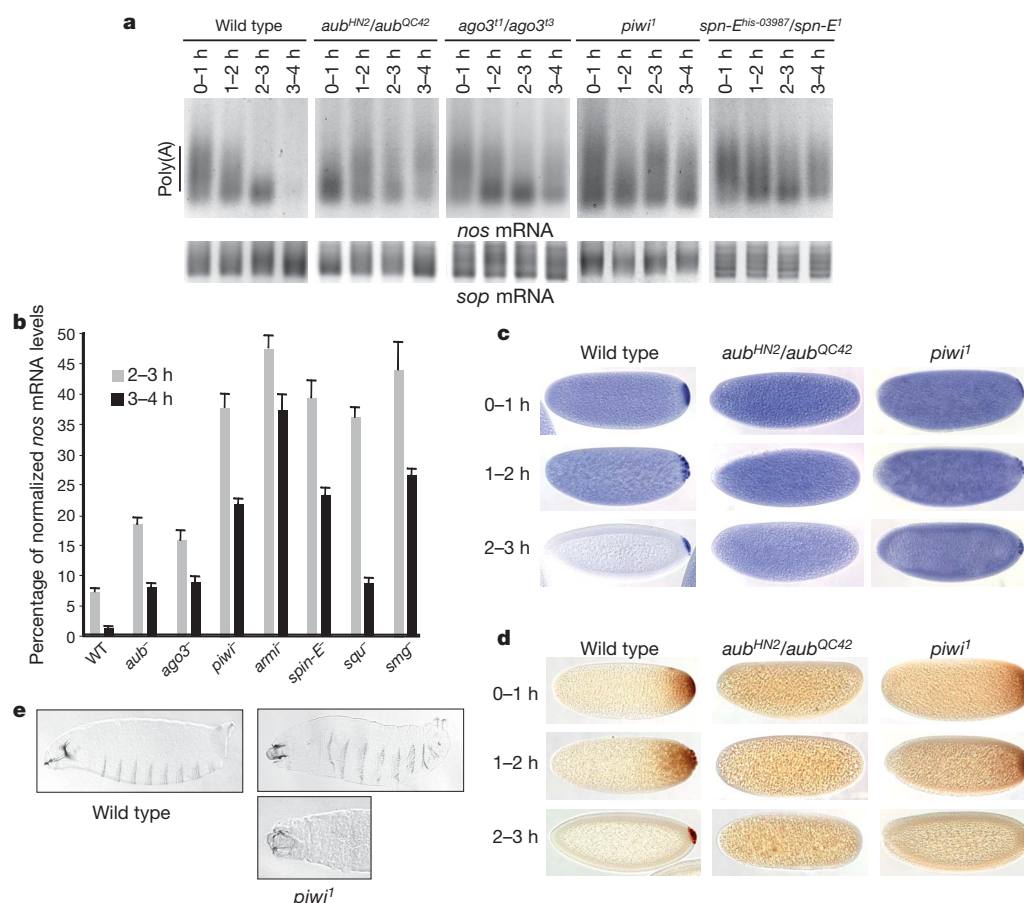
Figure 1 | The piRNA pathway is required for *nos* mRNA deadenylation and decay as well as translational repression in the bulk cytoplasm of the embryo. a, b, Poly(A) test assays and RT–qPCR of *nos* mRNA. Mutant females of the indicated genotypes were crossed with wild-type males. The *sop* mRNA was used as a control in **a**. **b,** Levels of *nos* mRNA in 2–3-h and 3–4-h embryos. WT, wild type. Mean value of three quantifications, error bars correspond to standard deviation (s.d.). **c,** *In situ* hybridizations of *nos* mRNA. **d,** Immunostaining of embryos with anti-Nos antibody. **e,** Cuticle preparations of *piwi[1]* embryos showing head defects (rudimentary head skeleton (top), head skeleton replaced by a hole (bottom)). **c–e,** Magnification of images is ×20. Two per cent of embryos from *piwi[1]* germline clones produced a cuticle (*n* = 1,060); among those, 22/23 had head defects. No embryos from *aub[N11]*/*aub[HN2]* (*n* = 1,230) or *aub[QC42]*/*aub[HN2]* (*n* = 813) females produced a cuticle.

colocalized with Smg and CCR4 in the bulk of syncytial embryos, in both cytoplasmic foci and a diffusely distributed cytoplasmic pool (Fig. 2b and Supplementary Fig. 6b). Importantly, the distributions of CCR4 and Smg depended on the piRNA pathway, as they were strongly affected in *aub* and *spn-E* mutant embryos. Although global amounts of CCR4 and Smg did not decrease in mutant embryos, CCR4 foci strongly increased in size, whereas Smg foci decreased in size or disappeared (Fig. 2c, d). This suggests that subsets of CCR4 and Smg foci have different functions and that deadenylation may take place diffusely in the cytoplasm. These results demonstrate a functional link between CCR4-mediated deadenylation and the piRNA pathway.

Co-immunoprecipitation experiments showed that Aub co-precipitated Smg, CCR4 and Ago3 in the absence of RNA, indicating the presence of these proteins in a common complex (Fig. 3a and Supplementary Fig. 7a, b). Smg also co-precipitated CCR4, Aub and Ago3 (Fig. 3b and Supplementary Fig. 7c); however, Piwi was not found to co-precipitate Smg or CCR4 (data not shown). Importantly, Smg, CCR4 and Ago3 also co-precipitated with Aub in *osk[54]* mutant embryos that are defective in pole plasm assembly[27], indicating the presence of this complex outside the pole plasm (Fig. 3a). Next we showed that *nos* mRNA co-precipitated with Aub in both wild-type and *osk[54]* embryos. The amount of *nos* mRNA was similar in Aub and Smg immunoprecipitates (Fig. 3c).

These findings show that the Argonaute proteins Aub and Ago3 associate with Smg and the CCR4 deadenylase complex to directly regulate *nos* mRNA in the bulk cytoplasm of early embryos.

The *nos* 3′ untranslated region (UTR) contains Smg-binding sites located in its 5′-most region (referred to as the translational control element (TCE))[16]. We searched for piRNAs sequenced from early embryos and presumed capable of targeting *nos* 3′ UTR based on their sequence complementarity. Notably, a specific region located in the 3′-most part of the 3′ UTR could be targeted by over 200 copies of piRNAs originating from two transposable elements, *412* and *roo* (Fig. 4a and Supplementary Fig. 8). piRNAs complementary to *nos* 3′ UTR were visualized by northern blots. In addition, piRNAs predicted to target *nos* 3′ UTR co-immunoprecipitated with Aub (Fig. 4b). We used *nos* genomic transgenes deleted for different parts of the 3′ UTR[16] to address the requirement of the corresponding regions for *nos* mRNA deadenylation. We have shown previously that the TCE (nucleotides 1–184) is required for *nos* mRNA poly(A) tail shortening, consistent with the role of Smg in this process[13]. Deletion of region 184–403 (*nos(Δ1)*) had no effect, whereas poly(A) tails from the transgene deleted for the region 403–618 (*nos(Δ2)*) were elongated in 3–4-h embryos (Fig. 4c and Supplementary Fig. 12). This could indicate regulation by the miRNA predicted by miRBase to target this region. Deletion of 618–844 in the *nos* 3′ UTR (*nos(Δ3)*) had a strong effect on *nos* deadenylation (Fig. 4c and Supplementary Figs 9 and 12). Consistent with this, *nos* mRNA levels produced by this transgene remained mostly stable (Fig. 4d). This resulted in defects in embryo patterning: a total of 35% (*n* = 1,894) of embryos from *nos(Δ3)* females did not hatch and among them 86% (*n* = 28) showed head skeleton defects (Fig. 4e). We next deleted specific sequences complementary to
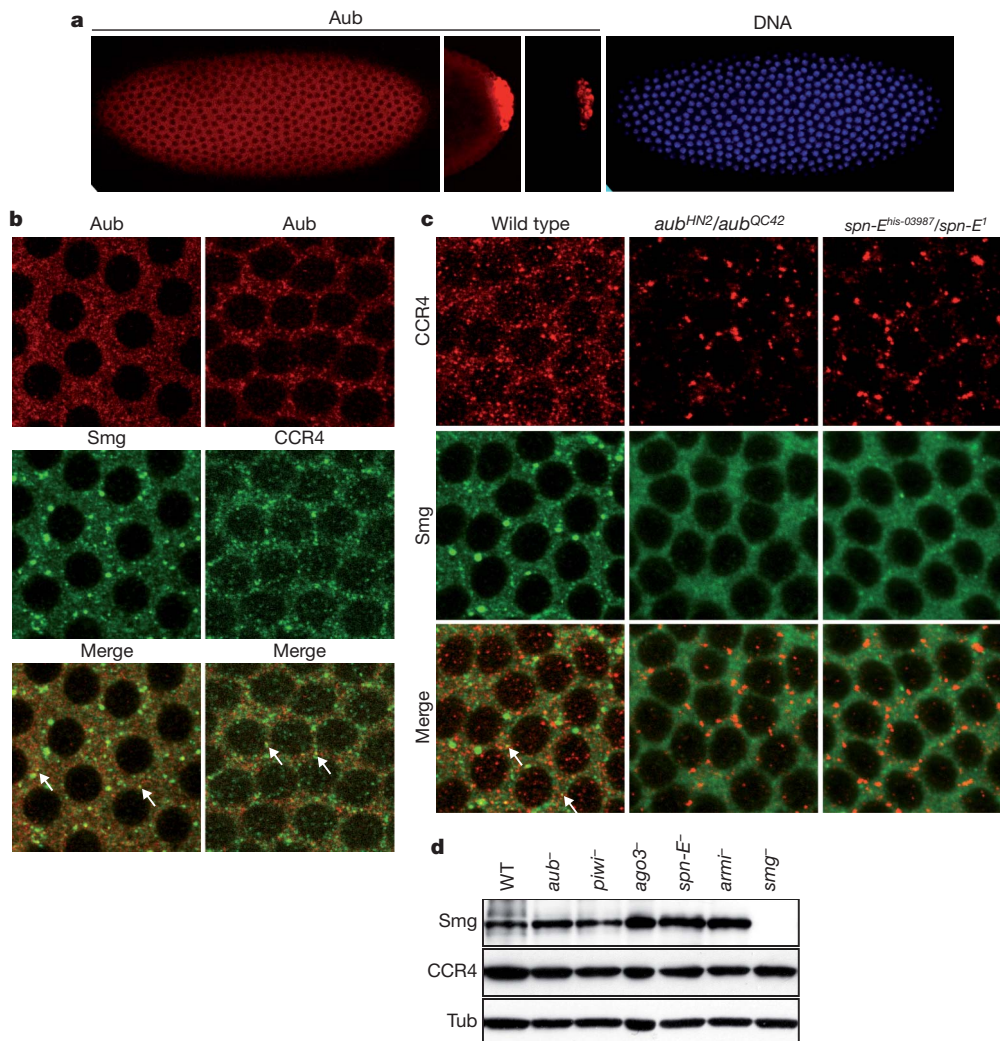
Figure 2 | Aub is present in the bulk of the embryo and the piRNA pathway is required for CCR4 and Smg cytoplasmic distributions.
a, Confocal images of cytoplasmic expression of Aub in the embryo. Syncytial blastoderm embryo at nuclear cycle 11, anterior is to the left. Pole cells of the same embryo, at the same setting (middle Aub panel) and at lower intensity (right Aub panel)[20,25]. 4′,6-Diamidino-2-phenylindole (DAPI) staining (right panel). Magnification, ×20.
b, Double immunostaining of embryos at nuclear cycles 11/12 with anti-Aub and anti-Smg, or anti-Aub and anti-CCR4. Arrows indicate examples of small foci showing colocalization in **b** and **c** (magnification, ×100). **c**, Smg and CCR4 cytoplasmic distributions are affected in *aub* and *spn-E* mutant embryos. Double immunostaining of embryos at nuclear cycle 11 with anti-CCR4 and anti-Smg. **d**, Western blots of proteins from 0–2-h embryos revealed with anti-Smg and anti-CCR4. α-Tubulin (Tub) was used as a loading control.

412 (15 nucleotides) and *roo* (11 nucleotides) retrotransposon piRNAs (Supplementary Fig. 8). These short deletions, either independently or in combination, affected *nos* mRNA deadenylation (Fig. 4f and Supplementary Fig. 12).
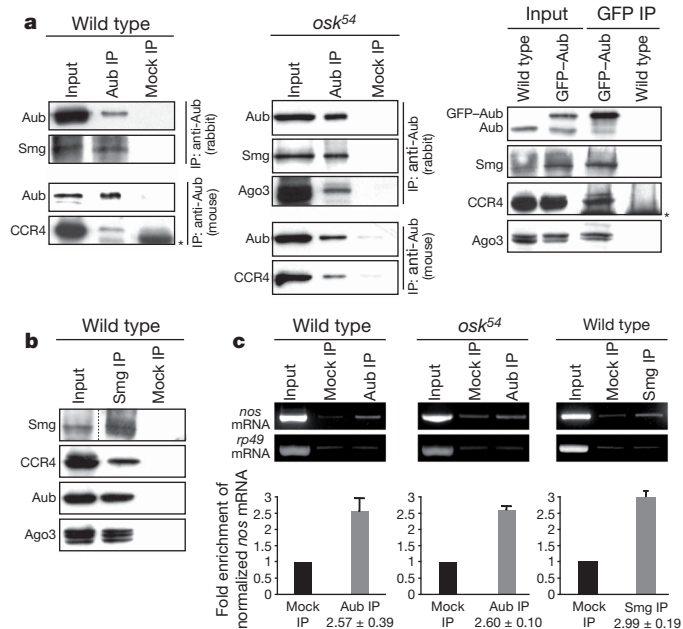
To support further the role of retrotransposon piRNAs in *nos* mRNA regulation, we blocked *412* and *roo* piRNAs by injecting specific 2′-O-methyl anti-piRNA in embryos[28], and recorded cuticles as a functional assay of Nos ectopic synthesis at the anterior pole. Injection of anti-piRNA(*412*) or anti-piRNA(*roo*) resulted in specific head development defects (Fig. 4g).

Together, these results provide strong evidence that an interaction between piRNAs and *nos* mRNA is required for *nos* mRNA deadenylation and translational repression in the first hours of embryogenesis.

We have identified a new function of the piRNA pathway in the regulation of maternal mRNAs. Recently, piRNAs derived from the 3′ UTRs of cellular transcripts have been identified in gonadal somatic cells, although their biological role has not been clarified[29,30]. Here we propose that piRNAs, in complex with Piwi-type Argonaute proteins Aub and Ago3, target *nos* maternal mRNAs and recruit or stabilize the CCR4–NOT deadenylation complex together with Smg (Supplementary

Figure 3 | Aub, Ago3, Smg, CCR4 and *nos* mRNA are present in a common complex in the bulk of the embryo. a, Co-immunoprecipitations of Smg, CCR4 and Ago3 with Aub in 0–2-h embryo extracts. Anti-Aub and anti-green fluorescent protein (GFP) were used for immunoprecipitations (IP) in wild-type, *osk[54]* and GFP–Aub-expressing embryos, respectively. The asterisks indicate immunoglobulins. **b**, Co-immunoprecipitations of CCR4, Aub and Ago3 with Smg in 0–2-h wild-type embryo extracts. **c**, Quantification of *nos* mRNA enrichment in Aub and Smg immunoprecipitations. Extracts from 0–2-h wild-type or *osk[54]* embryos were immunoprecipitated with anti-Aub (rabbit), or anti-Smg. For quantifications performed by RT–qPCR, the ratio of *nos* mRNA/*rp49* mRNA was set to 1 in the mock immunoprecipitation. Mean value of three quantifications, error bars correspond to s.d. *rp49* was used as a control mRNA.
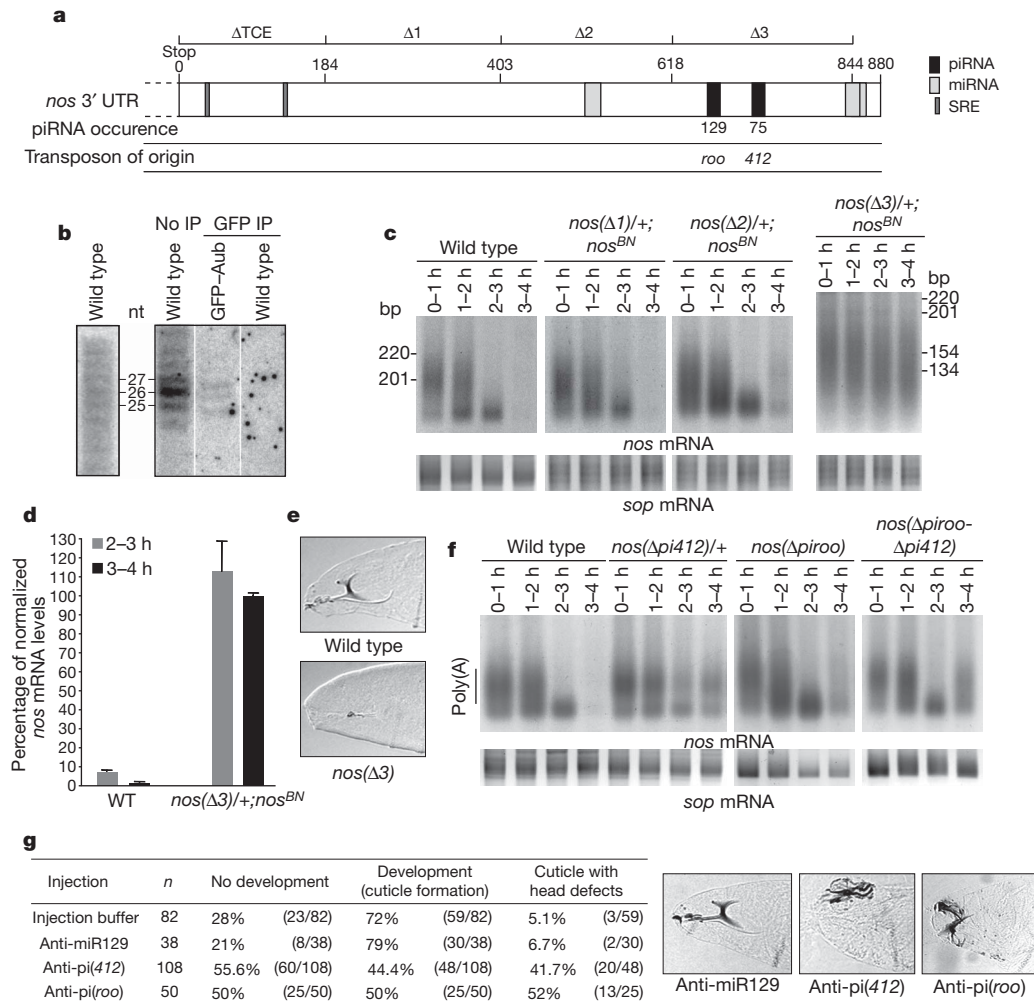
Figure 4 | piRNAs target a specific region in *nos* 3′ UTR that is required for *nos* mRNA deadenylation. a, Schematic representation of the *nos* 3′ UTR. The regions deleted in *nos* genomic transgenes (*nos(Δ)*) are indicated on the upper line[16]. The Smg recognition elements (SRE) and the piRNA and miRNA target sites are indicated. Predictions of miRNA-targeted regions are from miRBase (miR-31a, miR-314 and miR-263b from proximal to distal). piRNA occurrences in the data sets[19,20] are indicated. **b**, Northern blots of 0–2-h embryos probed with riboprobes corresponding to the sense *nos* 3′ UTR (position 403–844) (left) and to the antisense *412* piRNA (right). Anti-GFP immunoprecipitations (IP) were performed using wild-type and GFP–Aub-expressing embryos. **c**, *nos* poly(A) test assays. For *nos(Δ3)* the fragment amplified in the poly(A) test is shorter than the fragment amplified in the other

nos poly(A) tests (Supplementary Fig. 9). bp, base pairs; nt, nucleotides. **d**, Quantification of *nos* mRNA levels from the *nos(Δ3)* transgene by RT–qPCR. Mean value of three quantifications, error bars correspond to s.d. **e**, Cuticle preparations of embryos from *nos(Δ3)* females (lack of head skeleton). **e, g**, Magnification, ×20. **f**, *nos* poly(A) tests from embryos containing *nos* genomic transgenes in which sequences complementary to *412* piRNA, *roo* piRNA, or both sequences have been deleted. The *sop* mRNA was used as a control in **c** and **f. g**, Injection of 2′-O-methyl anti-piRNA in embryos. Control injections were with injection buffer alone or with the irrelevant anti-miR129. Examples of cuticles following injections of anti-miR129 (wild-type head skeleton), anti-pi(*412*) and anti-pi(*roo*) (affected head skeleton).

Fig. 10). These interactions induce rapid mRNA deadenylation and decay. Thus, activation of mRNA deadenylation represents a new direct mechanism of action for the piRNA pathway with an essential developmental function during the first steps of embryogenesis.

Smg is a general factor for mRNA decay during early embryogenesis[12]. Because Aub and Ago3 are present in a complex with Smg in early embryos, a proportion of Smg mRNA targets could be regulated by the piRNA pathway. Consistent with this, other maternal mRNAs that are destabilized during early embryogenesis are targeted by abundant piRNAs and their deadenylation depends on the piRNA pathway (Supplementary Fig. 11).

These piRNAs involved in gene regulation are generated from transposable element sequences. Although transposable elements have been described to be essential for genome dynamics and evolution, their immediate function within an organism has remained rather elusive. This study provides evidence for a co-evolution between transposable elements and the host genome and reveals the direct developmental

function of transposable elements in embryonic patterning, through the regulation of gene expression.

METHODS SUMMARY

RNA and proteins were manipulated using methods described previously and reported in Methods.

Embryo injections. Injections of embryos were performed laterally with 400 μM of 2′-O-methyl oligonucleotides as reported previously[28]. The injection buffer was 0.5 mM NaPO₄, 5 mM KCl. Sequences of 2′-O-methyl oligonucleotides are indicated in Methods.

Bioinformatics. A total of 29,108,987 piRNAs sequenced from 0–1-h embryos (GSM286613 and GSM286604 data sets[19]) and from 0–2-h embryos (GSM327625, GSM327626, GSM327627, GSM327628 and GSM327629 data sets[20]) was blasted against *nos* 3′ UTR using the following parameters: a National Center for Biotechnology Information (NCBI) blast with an E value of 100 and a 14-nucleotide match and a Washington University (WU)-blast with an E value of 10 and an 11-nucleotide match. Regions potentially targeted by piRNAs with an occurrence of less than ten were not considered.

1. Saito, K. *et al.* Specific association of Piwi with rasiRNAs derived from retrotransposon and heterochromatic regions in the *Drosophila* genome. *Genes Dev.* **20,** 2214–2222 (2006).
2. Vagin, V. V. *et al.* A distinct small RNA pathway silences selfish genetic elements in the germline. *Science* **313,** 320–324 (2006).
3. Brennecke, J. *et al.* Discrete small RNA-generating loci as master regulators of transposon activity in *Drosophila. Cell* **128,** 1089–1103 (2007).
4. Yin, H. & Lin, H. An epigenetic activation role of Piwi and a Piwi-associated piRNA in *Drosophila melanogaster. Nature* **450,** 304–308 (2007).
5. Aravin, A. A. *et al.* A piRNA pathway primed by individual transposons is linked to *de novo* DNA methylation in mice. *Mol. Cell* **31,** 785–799 (2008).
6. Lim, A. K., Tao, L. & Kai, T. piRNAs mediate posttranscriptional retroelement silencing and localization to pi-bodies in the *Drosophila* germline. *J. Cell Biol.* **186,** 333–342 (2009).
7. Cox, D. N. *et al.* A novel class of evolutionarily conserved genes defined by *piwi* are essential for stem cell self-renewal. *Genes Dev.* **12,** 3715–3727 (1998).
8. Chen, Y., Pane, A. & Schupbach, T. *cutoff* and *aubergine* mutations result in retrotransposon upregulation and checkpoint activation in *Drosophila. Curr. Biol.* **17,** 637–642 (2007).
9. Klattenhoff, C. *et al. Drosophila* rasiRNA pathway mutations disrupt embryonic axis specification through activation of an ATR/Chk2 DNA damage response. *Dev. Cell* **12,** 45–55 (2007).
10. Pane, A., Wehr, K. & Schupbach, T. *zucchini* and *squash* encode two putative nucleases required for rasiRNA production in the *Drosophila* germline. *Dev. Cell* **12,** 851–862 (2007).
11. Semotok, J. L. *et al.* Smaug recruits the CCR4/POP2/NOT deadenylase complex to trigger maternal transcript localization in the early *Drosophila* embryo. *Curr. Biol.* **15,** 284–294 (2005).
12. Tadros, W. *et al.* SMAUG is a major regulator of maternal mRNA destabilization in *Drosophila* and its translation is activated by the PAN GU kinase. *Dev. Cell* **12,** 143–155 (2007).
13. Zaessinger, S., Busseau, I. & Simonelig, M. Oskar allows *nanos* mRNA translation in *Drosophila* embryos by preventing its deadenylation by Smaug/CCR4. *Development* **133,** 4573–4583 (2006).
14. Bushati, N., Stark, A., Brennecke, J. & Cohen, S. M. Temporal reciprocity of miRNAs and their targets during the maternal-to-zygotic transition in *Drosophila. Curr. Biol.* **18,** 501–506 (2008).
15. Gavis, E. R. & Lehmann, R. Translational regulation of *nanos* by RNA localization. *Nature* **369,** 315–318 (1994).
16. Dahanukar, A. & Wharton, R. P. The Nanos gradient in *Drosophila* embryos is generated by translational regulation. *Genes Dev.* **10,** 2610–2621 (1996).
17. Dahanukar, A., Walker, J. A. & Wharton, R. P. Smaug, a novel RNA-binding protein that operates a translational switch in *Drosophila. Mol. Cell* **4,** 209–218 (1999).
18. Giraldez, A. J. *et al.* Zebrafish MiR-430 promotes deadenylation and clearance of maternal mRNAs. *Science* **312,** 75–79 (2006).
19. Chung, W. J., Okamura, K., Martin, R. & Lai, E. C. Endogenous RNA interference provides a somatic defense against *Drosophila* transposons. *Curr. Biol.* **18,** 795–802 (2008).
20. Brennecke, J. *et al.* An epigenetic role for maternally inherited piRNAs in transposon silencing. *Science* **322,** 1387–1392 (2008).
21. Gunawardane, L. S. *et al.* A slicer-mediated mechanism for repeat-associated siRNA 5′ end formation in *Drosophila. Science* **315,** 1587–1590 (2007).
22. Li, C. *et al.* Collapse of germline piRNAs in the absence of Argonaute3 reveals somatic piRNAs in flies. *Cell* **137,** 509–521 (2009).
23. Cook, H. A., Koppetsch, B. S., Wu, J. & Theurkauf, W. E. The *Drosophila* SDE3 homolog *armitage* is required for *oskar* mRNA silencing and embryonic axis specification. *Cell* **116,** 817–829 (2004).
24. Malone, C. D. *et al.* Specialized piRNA pathways act in germline and somatic tissues of the *Drosophila* ovary. *Cell* **137,** 522–535 (2009).
25. Harris, A. N. & Macdonald, P. M. Aubergine encodes a *Drosophila* polar granule component required for pole cell formation and related to eIF2C. *Development* **128,** 2823–2832 (2001).
26. Megosh, H. B., Cox, D. N., Campbell, C. & Lin, H. The role of PIWI and the miRNA machinery in *Drosophila* germline determination. *Curr. Biol.* **16,** 1884–1894 (2006).
27. Lehmann, R. & Nusslein-Volhard, C. The maternal gene nanos has a central role in posterior pattern formation of the *Drosophila* embryo. *Development* **112,** 679–691 (1991).
28. Leaman, D. *et al.* Antisense-mediated depletion reveals essential and specific functions of microRNAs in *Drosophila* development. *Cell* **121,** 1097–1108 (2005).
29. Saito, K. *et al.* A regulatory circuit for *piwi* by the large Maf gene *traffic jam* in *Drosophila. Nature* **461,** 1296–1299 (2009).
30. Robine, N. *et al.* A broadly conserved pathway generates 3′UTR-directed primary piRNAs. *Curr. Biol.* **19,** 2066–2076 (2009).

Author Contributions C.R. and C.P. designed and performed the experiments, analysed the data and contributed equally to the study. A.-C.M. contributed to the generation of DNA constructs, B.F. contributed to poly(A) test assays in Fig. 1. A.B., N.R. and E.C.L. performed the bioinformatic analyses. A.P. performed northern blots. M.S. designed the study, analysed data and wrote the paper. All authors discussed the results and commented on the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to M.S. (Martine.Simonelig@igh.cnrs.fr).

METHODS

***Drosophila* stocks and genetics.** The w^{1118} stock was used as a control. Mutant stocks were: *w; Sp aub^{N11} bw/CyO, aub^{HN2} cn^1 bw^1/CyO, aub^{QC42} cn^1 bw^1/CyO* (ref. 31), *mnk^{P6}, mnk^{P6} aub^{HN2}/CyO, mnk^{P6} aub^{QC42}/CyO* (ref. 9), *mnk^{P6}; armi^1/ SM6-TM6B, mnk^{P6}; armi$^{72.1}$/SM6-TM6B* (ref. 9), *piwi^1 neo-FRT40A/CyO* (ref. 7), *spn-E^1/TM3, ry^{506} spn-E$^{hls-03987}$ e/TM3* (ref. 32), *armi^1/TM3, armi$^{72.1}$/TM3* (ref. 23), *cn^1 bw^1 squ^{HE47}/CyO, cn^1 bw^1 squ^{PP32}/CyO* (ref. 10), *bw; st ago3^{t1}/TM6B Tb, bw; st ago3^{t3}/TM6B Tb* (ref. 22), *osk^{54} spn-E^1/TM3, osk^{54} spn-E$^{hls-03987}$ e/TM3*. *piwi^1* mutant embryos are from germline clones that were induced with two 1.5-h heat shocks at 37 °C during the second- and third-instar larval stages, using the flippase-dominant female sterile technique[33]. *smg* mutants were *smg^1* and a deficiency overlapping *smg, Df(Scf^{R6})* (ref. 17). The *nos^{BN}* mutant does not produce *nos* mRNA[34]. *osk^{54}* is a null allele. GFP–Aub was expressed following crosses between the germline driver *nos-Gal4:VP16* (ref. 35) and *UASp-GFP-Aub* (ref. 25). *nos(Δ)* stocks are transgenic lines containing a *nos* genomic transgene in which different parts of the 3′ UTR have been removed[16]. *nos(Δ)* stocks were a gift from R. Wharton. The *nos(Δpi412)* and *nos(Δpiroo)* transgenes, in which 15 nucleotides (TATATTTATTCAATT) and 11 nucleotides (AACACACATAT) were deleted, respectively, were generated as follows. The pBSKS-R5561 (containing a 5.7-kb *nos* genomic fragment, a gift from R. Wharton) was used as a template for PCR reactions to produce the deletion. For each construct, two PCR reactions were performed using the following primers: for *nos(Δpi412)*, 5′-CATTCCGATC AAAGCTGGGTTAACC (primer 1) and 5′-AAATTGATCAATGGTAAACAA TAACATATATATAT, which contains the 15-nucleotide deletion; and 5′-TA TATATATATATATATGTTATTGTTTACCATTGATCAATTT, which contains the 15-nucleotide deletion, and 5′-CTCCACCGCGGTGGCGGCCGC (primer 2). For *nos(Δpiroo)*, primer 1 and 5′-TATATATATATATATATATAGGAAATGAA TACTTGCGATACA, which contains the 11-nucleotide deletion; and 5′-TG TATCGCAAGTATTCATTTCCTATATATATATATATATATA, which contains the 11-nucleotide deletion and primer 2. For each construct, the two PCR products were annealed and used as a template for a third PCR reaction using primers 1 and 2. This third PCR product containing either the 15-nucleotide or the 11-nucleotide deletion surrounded by the restriction sites BglII and NotI was cloned into the TAcloning vector (pCRII) (Invitrogen) and sequenced. For the *nos(Δpiroo-Δpi412)* transgene, the PCR generating the *Δpiroo* deletion was done using pCRII containing the *nosΔpi412* deletion as a template. The BglII–NotI fragment containing the deletion was used to replace the BglII–NotI fragment in the original pBSKS-R5561. An EcoRI–NotI fragment containing the whole genomic fragment with the deletion was cloned into the pCaSpeR4. Transformant stocks were produced by BestGene.

Embryo injections. Sequences of 2′-O-methyl oligonucleotides were as follows. anti-pi(*412*), UCGGGCUGACAUAUAUUUAUUCAAUU; anti-pi(*roo*), UCCA AACACACAUAUAUAUAUAAAUA; anti-miR129-1, GCAAGCCCAGACCG CAAAAAG (human miR129-1 is not conserved in *Drosophila*).

RNA. Poly(A) test assays, RT–PCR and RT–qPCR were performed as described previously[13,36], and were made from two-to-four independent RNA preparations. For the *nos(Δ3)* transgene, a different *nos* specific primer was used (5′-GTC GTCGGCTACGCATTCATTGT), as the region normally amplified in *nos* poly(A) test assays is deleted in this transgene. We verified by sequencing that the poly(A) site used in mRNA from this transgene is identical to the one used in *nos* endogenous mRNA. Real-time PCR (qPCR) was performed with the

LightCycler System (Roche Molecular Biochemical) using *rp49* as a control mRNA[13]. For quantification of *nos* mRNA in 2–3-h and 3–4-h embryos, the levels were normalized with the levels of *nos* mRNA in 0–1-h embryos that were set to 100% for each genotype. Northern blots were performed as described previously[37]. The sequence of the riboprobe specific to *412* piRNA was 5′-GGGCUGAC AUAUAUUUAUUCAAUU.

RNA *in situ* hybridization and cuticle preparations. Whole-mount *in situ* hybridization and cuticle preparations were performed by standard methods. The probe for *in situ* hybridization was an antisense RNA made from the pN5 *nos* complementary DNA clone.

Antibodies, western blots, immunostaining and immunoprecipitations. Immunoprecipitations were performed as described previously[13] using 0–2-h embryos, and mouse anti-Aub (4D10 (ref. 21)), mock immunoprecipitations: mouse anti-haemagglutinin (12CA5 Developmental Studies Hybridoma Bank, for wild-type embryos) and mouse IgG (sc-2025 Santa Cruz Biotechnology, for *osk^{54}* embryos); rabbit anti-Aub (Abcam, ab17724), mock immunoprecipitations: rabbit IgG (sc-2027 Santa Cruz Biotechnology); mouse anti-GFP (monoclonal antibody 3E6 Invitrogen); guinea pig anti-Smg (gift from C. Smibert), mock immuno-precipitations: guinea pig pre-immune serum. Protein co-immunoprecipitations were performed in the presence of 0.1 μg μl^{-1} RNase A. Western blots and immuno-staining were performed as reported[38,39]. Antibodies for western blots were used at the following dilutions: guinea pig anti-Smg 1:5,000, anti-CCR4 1:1,000[40], anti-Piwi 1:20 (P4D2 (ref. 1)), anti-Aub 1:1,500 (4D10 (ref. 21)) and anti-Ago3 1:500 (9G3 (ref. 21)). Antibodies for immunostaining were used at the following dilutions: rabbit anti-Nos 1:1,000 (gift from A. Nakamura), guinea pig anti-Smg 1:1,000, anti-CCR4 1:300, anti-Aub 1:1,500 (4D10), anti-Piwi 1:1 (P4D2) and anti-Ago3 1:300 (9G3).

31. Schupbach, T. & Wieschaus, E. Female sterile mutations on the second chromosome of *Drosophila melanogaster*. II. Mutations blocking oogenesis or altering egg morphology. *Genetics* **129,** 1119–1136 (1991).
32. Gillespie, D. E. & Berg, C. A. *homeless* is required for RNA localization in *Drosophila* oogenesis and encodes a new member of the DE-H family of RNA-dependent ATPases. *Genes Dev.* **9,** 2495–2508 (1995).
33. Chou, T. B. & Perrimon, N. The autosomal FLP-DFS technique for generating germline mosaics in *Drosophila melanogaster*. *Genetics* **144,** 1673–1679 (1996).
34. Wang, C., Dickinson, L. K. & Lehmann, R. Genetics of *nanos* localization in *Drosophila*. *Dev. Dyn.* **199,** 103–115 (1994).
35. Rorth, P. Gal4 in the *Drosophila* female germline. *Mech. Dev.* **78,** 113–118 (1998).
36. Juge, F., Zaessinger, S., Temme, C., Wahle, E. & Simonelig, M. Control of poly(A) polymerase level is essential to cytoplasmic polyadenylation and early development in *Drosophila*. *EMBO J.* **21,** 6603–6613 (2002).
37. Pelisson, A., Sarot, E., Payen-Groschene, G. & Bucheton, A. A novel repeat-associated small interfering RNA-mediated silencing pathway downregulates complementary sense gypsy transcripts in somatic cells of the *Drosophila* ovary. *J. Virol.* **81,** 1951–1960 (2007).
38. Benoit, B. *et al.* An essential cytoplasmic function for the nuclear poly(A) binding protein, PABP2, in poly(A) tail length control and early development in *Drosophila*. *Dev. Cell* **9,** 511–522 (2005).
39. Benoit, B. *et al.* The *Drosophila* poly(A)-binding protein II is ubiquitous throughout *Drosophila* development and has the same function in mRNA polyadenylation as its bovine homolog *in vitro*. *Nucleic Acids Res.* **27,** 3771–3778 (1999).
40. Temme, C., Zaessinger, S., Meyer, S., Simonelig, M. & Wahle, E. A complex containing the CCR4 and CAF1 proteins is involved in mRNA deadenylation in *Drosophila*. *EMBO J.* **23,** 2862–2871 (2004).

The search for association

The list of human genetic variations is expanding; but an understanding of how they contribute to disease is still patchy.

BY MONYA BAKER

Everyone carries dangerous genetic mutations. But only in the past five years or so have researchers begun to use genome-wide association studies (GWAS) to scour human genetic samples for the signals of individual variations. Typically, such studies assess hundreds of thousands of genetic variants in thousands of individuals sorted by traits: a certain height, perhaps, or asthma or obesity. Genetic variants that occur more frequently in one group than in another are subjected to stringent statistical analyses to determine whether associations between them and the traits are the result of biology or mere chance.

As of 1 October, an online catalogue of GWAS contained nearly 700 publications linking some 3,000 variants to about 150 traits. The list of traits begins with abdominal aortic aneurysm and ends with YKL-40, a protein used as a biomarker for cancer. Other GWAS have identified correlations between genetic variants and smoking behaviour, sleep duration and general self-reported health. The catalogue is growing swiftly: 9 out of 16 research articles in the October issue of *Nature Genetics* report GWAS.

In their current incarnation, GWAS are running into a problem of diminishing returns. By collecting ever-larger samples, researchers are able to find more and more variant–trait associations, but these tend to have smaller and smaller effects. In fact, small effect sizes have been a hallmark of GWAS ever since the studies began. Originally, researchers hoped to find associations in which people carrying one variant would be several times more likely to have a trait than those carrying another. Instead, the effects found have been much more modest. An analysis published in June 2010 (ref. 1) pooled findings from several published GWAS that had each associated given traits with single nucleotide polymorphisms (SNPs) — the simplest and most common type of genetic variant, in which one DNA letter is changed to another. Extrapolating from previous findings, the researchers calculated that 201 SNPs associated with height could explain about 16% of genetic variance, 142 SNPs associated with Crohn's disease could explain about 20%, and 67 SNPs could explain about 17% of genetic variance in each of three common cancers.

Although genetic variants with small effects can still help to uncover fundamental biology with therapeutic implications, researchers hunting for those with larger effects are pinning their hopes on several advances: an onslaught of newly discovered simple polymorphisms, the ability to assess more complicated variants (see 'The tough new variants') and multiplying applications of sequencing. If the human genome were an archaeological site, these options would be equivalent to canvassing continents with metal detectors of varying convenience and reliability, or picking a handful of sites for a full excavation.

RARER SNPS, BIGGER EFFECTS?

GWAS are only as good as the SNPs they sample. Rather than directly finding mutations responsible for an effect, the standard technique identifies SNPs that tend to co-occur with it. And the SNPs that have already been profiled in GWAS may be the ones least likely to be linked with large effects. Because the most common variants were the first to be catalogued, they were also the first that vendors put on genotyping microarrays. To make sure that SNP microarrays could identify variants in the greatest possible number of samples, vendors chose variants that occurred across several geographic populations. These tended to be the SNPs that evolved first, so natural selection has had time to weed out harmful mutations that might have occurred nearby in the genome.

But multinational projects are now discovering and characterizing younger, rarer genetic variants. The 1000 Genomes Project aims to sequence 2,500 individuals, who represent an equal distribution from the continental regions of Africa, the Americas, Europe, and east and south Asia. The goal is to identify most of the variants that exist at frequencies of 1% or more in each of the populations studied, says David Altshuler, the project's co-leader and a human geneticist at the Broad Institute in Cambridge, Massachusetts. Similarly, the International HapMap Project has identified millions of SNPs and characterized their occurrence across populations. Sequencing data from ten geographic populations indicated that more than half of the
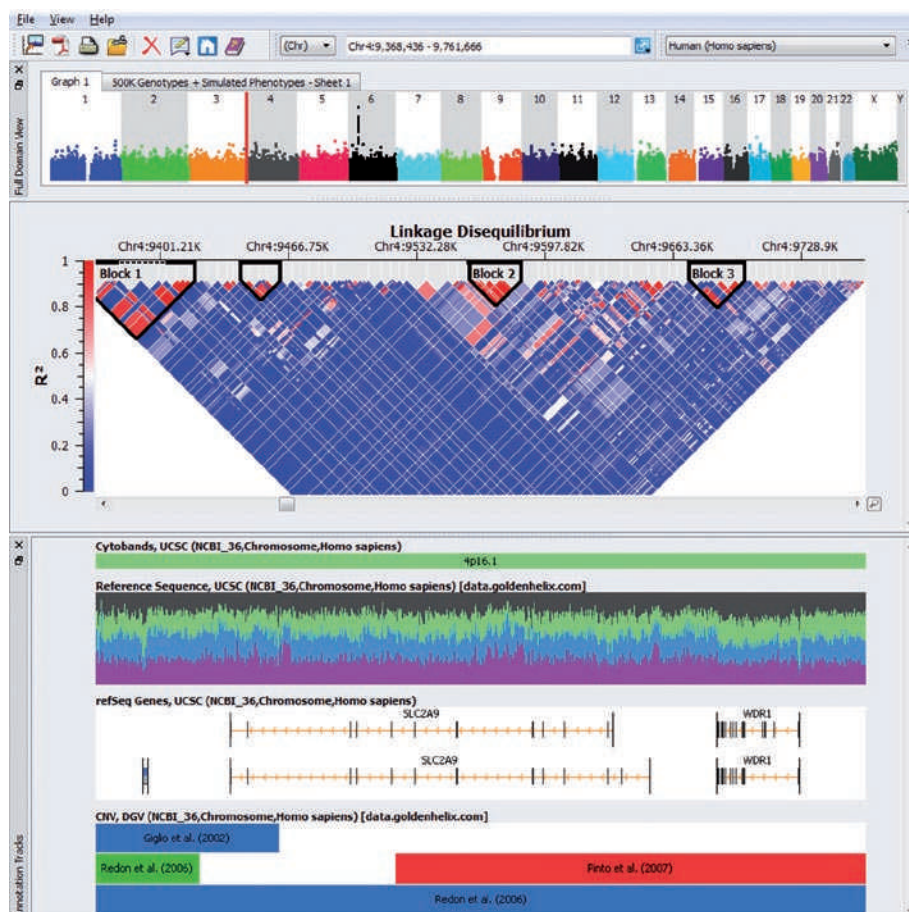
GOLDEN HELIX

Genome-wide association studies require several forms of statistical analyses.

L. RUBENSTEIN/BROAD INST.

identified genetic variants occur at frequencies of less than 5 per cent. More than one-third of newly discovered SNPs with frequencies of less than 0.5% were observed in only one population. Such discoveries mean that many more variants can be added to microarrays for assay, and so tested in GWAS, says David Bentley, chief scientific officer at Illumina, a genetics company in San Diego, California. "There is a new generation of GWAS that are fundamentally different from previous studies, because they capture a new fraction of variations that have previously been uncharted," he says.

Illumina and other commercial vendors have been modifying their microarrays in response to releases of data. Illumina unveiled its HumanOmni2.5-Quad DNA Analysis BeadChip in June this year — letting researchers assay 2.5 million SNPs and other variants — and plans to launch the Omni5 next year, for 5 million SNPs. Using the Omni5, researchers will be able to combine one set of comprehensive SNPs with specialized sets tuned to emerging sequencing data. Illumina's competitor Affymetrix, in Santa Clara, California, has in its catalogue products geared towards Chinese, Japanese, European and African ethnicities. A new microarray design allows researchers to design custom arrays containing 50,000 up to a planned 5 million SNPs using a database

David Altshuler: no one approach can explain heritability.

stocked with proprietary and public SNP data.

Nonetheless, it is not clear how effective adding to the available SNPs from healthy populations is going to be in finding SNPs associated with disease, says Christophe Lambert, chief executive of Golden Helix, a genetic-analysis company in Bozeman, Montana. This year, his company worked on an association study for Alzheimer's disease that failed to detect a signal from a variant known to boost risk for the condition. The variant, in the gene *APOE*, wasn't included on the commercial assay used in the test. Although a custom-designed array found the variant's association with the disease to be extremely significant ($P < 10^{-60}$), the standard array did not pick up its signal. "None of the SNPs on the standard chip was correlated strongly enough with the risk variant to detect it," says Lambert. Even when Lambert's team used data from the 1000 Genomes Project to 'impute' the presence of one SNP by detecting another, the analysis did not pick up on the association. Sampling more individuals or using denser microarrays might have helped, but identifying variants in diseased individuals would produce the most-informative SNPs for genotyping across populations, says Lambert.

Still, the ability to look more deeply within populations has intriguing possibilities. In a study published this September[3], researchers at deCODE Genetics in Reykjavik found that the same SNP was associated with glaucoma risk in Chinese and Icelandic populations, but in the former it was much rarer and indicated a much higher risk. And if different susceptibility variants show up near the same gene in different

David Goldstein: you have to choose what to pursue.

DUKE UNIV.

populations, researchers will have independently implicated that genomic area in the disease.

Working across populations and with rarer variants can get complicated, says Augustine Kong, head of statistics at deCODE. SNPs specific to a particular population could be difficult to replicate, and the lower the frequency of an allele, the larger the number of samples needed to detect an association. However, if rarer SNPs have stronger effects, larger sample sizes might not be necessary. Researchers are keen to find out whether a substantial number of the new variants discovered by genome-mapping projects will be associated with large effects. "Before, we just didn't have the technology to interrogate these low-frequency variants comprehensively," he says. "It gives you chances that you didn't have before to make discoveries."

SEQUENCING STRAIGHT TO CAUSAL VARIANTS
Some experts think that it is time to skip array-based GWAS that find SNPs associated with causative variants, and to hunt for contributing variants directly. Mary-Claire King is a geneticist at the University of Washington in Seattle, whose work in family studies identified the breast-cancer genes *BRCA1* and *BRCA2*. She says that even the rarer variants discovered by the 1000 Genomes Project are unlikely to be highly associated with disease. New variants

The tough new variants

When single nucleotide polymorphism (SNP) studies failed to explain much of the heritability of diseases, researchers began pinning their hopes on a trickier source of variability: copy number variation (CNV). Whereas SNPs — changes of one DNA letter into another — are relatively easy for microarrays to detect and for databases to compile and sort, CNVs are a headache to identify and classify. Certain stretches of DNA are duplicated, inverted or repeated in some individuals and missing from others. "It's more complicated and the data will always be a little more dirty," says Stephen Scherer, director of the Centre for Applied Genomics at the Hospital for Sick Children in Toronto, Canada. In some cases, researchers can detect CNVs using microarrays designed for detecting SNPs. Others use products designed to identify CNVs directly, from companies such as Agilent Technologies in Santa Clara, California, and Roche Nimblegen in Madison, Wisconsin. One Agilent array, designed with the Wellcome Trust Case Control Consortium,

detects about 11,000 common CNVs.

Measuring whether a nucleotide at a particular spot is A or G is easier than detecting how many times a certain sequence occurs. That concerns Peter Donnelly, director of the Wellcome Trust Centre for Human Genetics in Oxford, UK. "Because there was a long history of GWA studies that didn't replicate, the field insists on strong criteria for declaring an association," he says. "Yet when it moves to CNVs, which are harder to measure, the standards the field requires are weaker."

The jury is out on how much CNVs matter for common diseases. A study this year[8] profiled 3,423 CNVs, or perhaps half of all those larger than 500 base pairs. It found that most not only don't explain much disease, but are also so closely associated with common SNPs that they've already been explored, albeit indirectly.

Scherer is not so sure. He was part of a team that resequenced a human genome and compared it to a reference. It found that the genome differed from the reference in only 0.1% of SNPs, but in 1.2% of CNVs. The

analysis indicated that up to one-quarter of CNVs are not associated with SNPs, and so are likely to be missed by SNP studies[9].

As with SNPs, larger effects may be found in rarer and harder-to-measure variants. Scherer has done studies showing that people with autism-spectrum disorders carry more rare CNVs than do controls. To be certain that the CNVs were correctly typed, he and his colleagues ran subsets of samples through calling algorithms that convert an instrument's signals into a sequence of base pairs, and used two platforms (by Illumina, of San Diego, California, and Agilent) to identify them[10].

Scherer says that many research groups are still learning about CNVs and don't fully realize the need to validate their data. "People are looking for low-hanging fruit; they see what they want to see and publish it," he says. The situation is improving, with the maturation of databases that collect diverse data on variation. "Now that we have much better data sets to compare to, it's becoming more accurate." **M.B.**

are literally born every generation, she says, so a frequency rate of even 0.5% means that a variant has persisted for a while. "The question, is how common can an allele become if there is selection against it and none for it? Not very," says King. She advocates using sequencing within large families to find and track alleles that are inherited along with disease. Results from the 1000 Genomes Project will be useful, she says, not for finding SNPs to pursue but for filtering out variants that are not truly rare.

For David Goldstein, director of the Center for Genome Variation at Duke University in Durham, North Carolina, the main limitation for SNP-based GWAS is that they usually don't allow identification of the precise causal variant that influences the trait, but instead implicate a genomic region within which the causal variants must reside. The priority for research now, he says, is to focus on identifying the precise variants that contribute to disease; doing so will provide much more information about the relevant biological processes.

Peter Donnelly: copy number variants are hard to assess.

This year[4], researchers reported the first use of massively parallel sequencing to identify the gene responsible for a Mendelian disease, one which is caused by mutations in a single gene. Researchers at the University of Washington took samples from just four individuals with the developmental disorder Miller syndrome, including two siblings, and sequenced all the coding regions of their genomes using a technique called whole-exome sequencing, which relies on genome-capture products from companies such as Agilent Technologies in Santa Clara, California, and Roche Nimblegen in Madison, Wisconsin, followed by next-generation sequencing. Although still labour-intensive, the method requires only about 5% as much sequencing as a whole genome. By filtering identified variants against publicly available SNPs and other human exome data, the researchers found that all four subjects carried previously unidentified mutations in a single gene involved in synthesizing nucleotides; follow-up studies in three further people with Miller syndrome revealed mutations in the same gene. Since then, those and other researchers have published a slew of papers tying
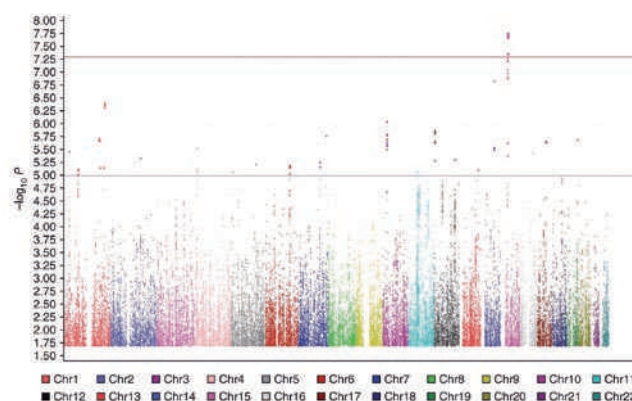
Plot showing an analysis of genetic variants' association with myopia. A locus on chromosome 15 has a *P* value of 10^{-14}, a very robust association.

a variety of other inherited conditions to variation in individual genes.

The signals for diseases caused by multiple genes will not be so clear, but bioinformatics techniques borrowed from SNP-based GWAS are being applied to sequencing data. To enable this, many types of variants must be placed into distinct categories so that they can be subjected to statistical analyses. Collapsing or binning methods count how many of a particular kind of variant are found in a gene or a predetermined stretch of base pairs, and then compare frequencies in individuals with the disease to those without it. But researchers are still learning how to weed out artefacts from the sequencing data, says Goldstein. With sequencing, "you have to pick which variants to pursue, and that's prone to statistical abuse", he says. "People could say the association is weak, but it makes sense." (See 'Seeing more SNPs'.)

One problem is that it is unclear how to classify different kinds of mutations. It might make sense, for example, to lump together different mutations in the same gene that stop translation early. But what about apparently silent mutations, or others whose effects on protein

1: rs8170
2: rs4808611
3: rs8100241
4: rs2363956
5: rs3745185

A: rs4808075
B: rs10419397
C: rs56069439
D: rs1465581
E: rs73511822
F: rs67397200
G: rs61494113
H: rs4808616
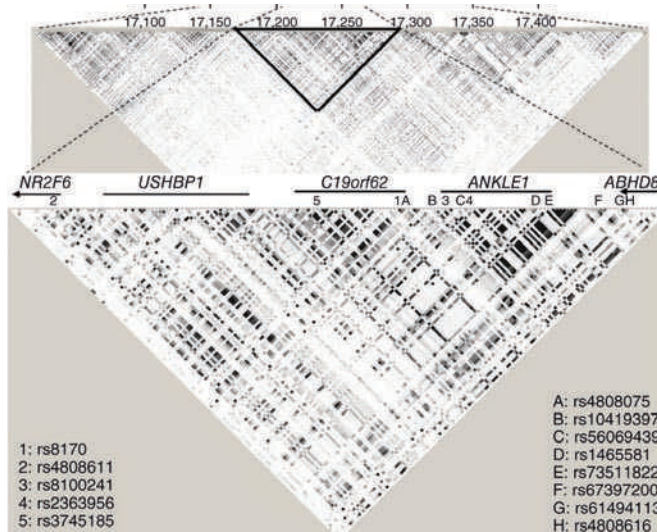
Studies must consider how variants are inherited together. Variants not directly measured (letters) can be imputed from those that are (numbers).

products seem minor? Another issue is working out the parameters for determining effective controls; because sequencing studies have smaller sets of controls, researchers need to be more rigorous to make sure that, for example, control sets don't include individuals with late-onset disease. But those problems won't send researchers back to SNP-based microarrays, says Goldstein. "If you were launching a new study now on a common disease, you'd turn to sequencing-based studies."

THE COMMON TOUCH

First-generation GWAS looked at common genetic variants, but many initial-sequencing studies concentrate on finding individuals with extreme forms of disease, because cost limits them to small sample sizes. Prices are changing quickly and vary by sample and technology, but at the moment, genotyping a sample for millions of SNPs costs about US$400, whereas sequencing an entire genome costs around $10,000. Finding the genetic basis for an extreme form of disease can shed light on more common forms: in the 1980s, well before sequencing, family-based studies of severe cholesteraemia led to the discovery of the low-density lipoprotein receptor gene, common variants of which are now associated with high cholesterol levels[5]. But Altshuler thinks that studies linking rare variants with strong effects need to be followed up to understand how the relevant genes contribute to common forms of the disease. Trying to study common diseases as if they were single-gene disorders will be of limited use, he says. "Mendelian forms of type 2 diabetes explain less than 1% of heritability, but GWAS explain about 10%," he says. To fully understand both the inheritance and mechanisms of common diseases, he says, it will be necessary to study the diseases as they occur in the general population.

Another problem with sequencing is that it is slow — one reason why Illumina's Bentley, whose company does both SNP genotyping and sequencing, says that he doesn't expect to see a decline in demand for microarrays any time soon. "With our best efforts at the moment we are sequencing one genome every three days; we can genotype more than 50 to 100 samples every three days," he says.

Even if GWAS continue to find only small effect sizes, they can still have a large impact, says Peter Donnelly, director of the Wellcome Trust Centre for Human Genetics in Oxford, UK. "There is real value in working out the genetic architecture of a disease, regardless of

Seeing more SNPs

As genome-wide association studies (GWAS) get larger, the technical challenges pile up, and an onslaught of dense microarrays is compounding the issue by encouraging researchers to combine data sets. Genotyping a few-dozen single nucleotide polymorphisms (SNPs) in a sample is not much cheaper than genotyping hundreds of thousands, says Peter Donnelly, director of the Wellcome Trust Centre for Human Genetics in Oxford, UK. So rather than designing a targeted follow-up study on a handful of SNPs, researchers are more likely to try to replicate an association through meta-analysis, using samples that have been fully genotyped elsewhere. "That needs care," says Donnelly. Even in straightforward GWAS, everything that looks like a signal is probably an artefact, he says. Combining results typed on one platform in one lab and on another in a different lab creates more opportunities for artefacts.

Even when cases and controls are processed by the same group, all the cases can be on one set of microarray plates and all the controls on another. This introduces potential for systemic error that sometimes leads to up to 30% of the data being discarded, says Christophe Lambert, chief executive of Golden Helix in Bozeman, Montana, which provides software and analytical services for genetic research. "Everyone is running these experiments and asking the statisticians to fix the problems, when a simple block randomization at the beginning could have fixed it."

Some problems occur before the sample is collected, says James Clough of Oxford Gene Technology, a genotyping-services firm. "Samples will be collected in multiple centres and multiple countries." That can pose challenges when clinical standards vary. The best studies put more effort into collecting phenotypes than collecting samples, he says.

Careful characterization of phenotype could make genetic signals more apparent, says Greg Gibson, director of the Center for Integrative Genomics at the Georgia Institute of Technology in Atlanta. Many aspects of phenotype are extremely variable, so longitudinal measurements of factors such as blood-lipid levels, body-mass index or toxin exposure could control for transient effects and effectively boost genetic signals. GWAS could be more successful at implicating genes if they concentrate on qualities more closely tied to genetics, such as lipid levels or endophenotypes, he says. "Just mapping genotype to disease is several steps away from gene expression." **M.B.**

what it turns out to be. For example, even if all the genetic components of a disease were based in very many common variants with small effects, it would be good to know that." And even if the effects of variants of a gene in a general population are small, those of modulating that gene with a drug can be large. For instance, variations in the gene encoding 3-hydroxy-3-methylglutaryl coenzyme A reductase have been connected in GWAS with small effects[6] on cholesterol levels, but the statin drugs that modulate that gene product are effective and very widely prescribed. Although statins were not inspired by GWAS, such studies have turned up surprising connections with therapeutic implications, such as the role of the immune system in age-related macular degeneration, or of cell-cycle regulators in type 2 diabetes. In fact, says Altshuler, such results could be useful for focusing sequencing studies. "The genome-wide association paradigm might be that you find the gene using GWAS, and then sequence to find the rarer variants."

One of the biggest GWAS so far assessed samples from more than 100,000 individuals for more than 2 million SNPs, and identified 95 loci associated with variation in cholesterol and triglyceride levels in blood, 59 of which had never been reported before, and many of which were not near genes known to be associated with lipid metabolism[5]. Follow-up experiments in mice not only showed that some newly implicated genes had direct effects on plasmid lipid levels, but also identified a new cell-signalling pathway that could be targeted for therapeutic intervention. Another study[7] examined four genes that had been implicated by GWAS as contributing to high blood-triglyceride levels. Common variants explained less than 10% of observed variation, so researchers sequenced the genes to identify rare missense and non-sense variants — two categories of mutations likely to change protein function. Nearly twice as many of these were found in affected individuals than in controls.

DIFFERENT STRATEGIES

The debate over the best approach for finding causal variants, says Altshuler, reflects researchers' various options for studying disease, and their limited funds. The decision whether to sequence a handful of samples or genotype thousands depends on whether researchers believe that a disease will be explained by a few rare variants or many common ones.

The answer will vary by disease. Current GWAS, for example, explain more heritability for autoimmune disorders and late-onset diseases such as Alzheimer's and heart disease than for mental conditions such as schizophrenia and autism. Natural selection suggests ready explanations, although they are hard to prove. Almost by definition, late-onset diseases tend to affect individuals in their post-reproductive years, and so are less likely to be selected against. And some genetic variants that contribute to one disease might actually be protective against others, and so could be favoured by natural selection. Genetic variants for sickle-cell anaemia, for example, can help to prevent carriers from contracting malaria, and there are hints that genes causing predisposition to some autoimmune diseases also confer resistance to infection.

In an effort to gather concrete evidence on which technologies are best suited to explaining the inheritance of common diseases, Altshuler has begun a study, with Mike Boehnke at the University of Michigan and Mark McCarthy at the University of Oxford, to compare the same population using several techniques. In this case, the study will compare what Altshuler calls "extremes of risk": subjects who are at high risk for diabetes because of their age and weight but do not have the disease will be compared with slimmer, younger subjects who have been diagnosed with it. Presumably, individuals in the first group will carry relatively more protective variants, whereas those in the latter will have more susceptibility variants. About 2,600 people will be genotyped for 5 million SNPs, and be submitted to whole-exome and whole-genome sequencing.

Altshuler says that the study should not only uncover important information about diabetes, but also offer empirical data to help researchers choose the most appropriate technology, or combination of technologies. "We want to know what each approach finds that the others don't," he says. "Right now, no one actually knows which one is going to apply to which disease. Investigators have to take different bets." ∎

Monya Baker *is technology editor for* Nature *and* Nature Methods.

1. Park, J. H. *et al. Nature Genet.* **42,** 570–575 (2010).
2. International HapMap 3 Consortium *Nature* **467,** 52–58 (2010).
3. Thorleifsson, G. *et al. Nature Genet.* **42,** 906–909 (2010).
4. Ng, S. B. *et al. Nature Genet.* **42,** 30–35 (2010).
5. Teslovich, T. M. *et al. Nature* **466,** 707–713 (2010).
6. Burkhardt, R. *et al. Arterioscler. Thromb. Vasc. Biol.* **28,** 2078–2084 (2008).
7. Johansen, C. T. *et al. Nature Genet.* **42,** 684–687 (2010).
8. Wellcome Trust Case Control Consortium *Nature* **464,** 713–720 (2010).
9. Pang, A. W. *et al. Genome Biol.* **11,** R52 (2010).
10. Pinto, D. *et al. Nature* **466,** 368–372 (2010).

identified genetic variants occur at frequencies of less than 5 per cent. More than one-third of newly discovered SNPs with frequencies of less than 0.5% were observed in only one population. Such discoveries mean that many more variants can be added to microarrays for assay, and so tested in GWAS, says David Bentley, chief scientific officer at Illumina, a genetics company in San Diego, California. "There is a new generation of GWAS that are fundamentally different from previous studies, because they capture a new fraction of variations that have previously been uncharted," he says.

Illumina and other commercial vendors have been modifying their microarrays in response to releases of data. Illumina unveiled its HumanOmni2.5-Quad DNA Analysis BeadChip in June this year — letting researchers assay 2.5 million SNPs and other variants — and plans to launch the Omni5 next year, for 5 million SNPs. Using the Omni5, researchers will be able to combine one set of comprehensive SNPs with specialized sets tuned to emerging sequencing data. Illumina's competitor Affymetrix, in Santa Clara, California, has in its catalogue products geared towards Chinese, Japanese, European and African ethnicities. A new microarray design allows researchers to design custom arrays containing 50,000 up to a planned 5 million SNPs using a database

David Altshuler: no one approach can explain heritability.

stocked with proprietary and public SNP data.

Nonetheless, it is not clear how effective adding to the available SNPs from healthy populations is going to be in finding SNPs associated with disease, says Christophe Lambert, chief executive of Golden Helix, a genetic-analysis company in Bozeman, Montana. This year, his company worked on an association study for Alzheimer's disease that failed to detect a signal from a variant known to boost risk for the condition. The variant, in the gene *APOE*, wasn't included on the commercial assay used in the test. Although a custom-designed array found the variant's association with the disease to be extremely significant ($P < 10^{-60}$), the standard array did not pick up its signal. "None of the SNPs on the standard chip was correlated strongly enough with the risk variant to detect it," says Lambert. Even when Lambert's team used data from the 1000 Genomes Project to 'impute' the presence of one SNP by detecting another, the analysis did not pick up on the association. Sampling more individuals or using denser microarrays might have helped, but identifying variants in diseased individuals would produce the most-informative SNPs for genotyping across populations, says Lambert.

Still, the ability to look more deeply within populations has intriguing possibilities. In a study published this September[3], researchers at deCODE Genetics in Reykjavik found that the same SNP was associated with glaucoma risk in Chinese and Icelandic populations, but in the former it was much rarer and indicated a much higher risk. And if different susceptibility variants show up near the same gene in different

David Goldstein: you have to choose what to pursue.

populations, researchers will have independently implicated that genomic area in the disease.

Working across populations and with rarer variants can get complicated, says Augustine Kong, head of statistics at deCODE. SNPs specific to a particular population could be difficult to replicate, and the lower the frequency of an allele, the larger the number of samples needed to detect an association. However, if rarer SNPs have stronger effects, larger sample sizes might not be necessary. Researchers are keen to find out whether a substantial number of the new variants discovered by genome-mapping projects will be associated with large effects. "Before, we just didn't have the technology to interrogate these low-frequency variants comprehensively," he says. "It gives you chances that you didn't have before to make discoveries."

SEQUENCING STRAIGHT TO CAUSAL VARIANTS

Some experts think that it is time to skip array-based GWAS that find SNPs associated with causative variants, and to hunt for contributing variants directly. Mary-Claire King is a geneticist at the University of Washington in Seattle, whose work in family studies identified the breast-cancer genes *BRCA1* and *BRCA2*. She says that even the rarer variants discovered by the 1000 Genomes Project are unlikely to be highly associated with disease. New variants

The tough new variants

When single nucleotide polymorphism (SNP) studies failed to explain much of the heritability of diseases, researchers began pinning their hopes on a trickier source of variability: copy number variation (CNV). Whereas SNPs — changes of one DNA letter into another — are relatively easy for microarrays to detect and for databases to compile and sort, CNVs are a headache to identify and classify. Certain stretches of DNA are duplicated, inverted or repeated in some individuals and missing from others. "It's more complicated and the data will always be a little more dirty," says Stephen Scherer, director of the Centre for Applied Genomics at the Hospital for Sick Children in Toronto, Canada. In some cases, researchers can detect CNVs using microarrays designed for detecting SNPs. Others use products designed to identify CNVs directly, from companies such as Agilent Technologies in Santa Clara, California, and Roche Nimblegen in Madison, Wisconsin. One Agilent array, designed with the Wellcome Trust Case Control Consortium,

detects about 11,000 common CNVs.

Measuring whether a nucleotide at a particular spot is A or G is easier than detecting how many times a certain sequence occurs. That concerns Peter Donnelly, director of the Wellcome Trust Centre for Human Genetics in Oxford, UK. "Because there was a long history of GWA studies that didn't replicate, the field insists on strong criteria for declaring an association," he says. "Yet when it moves to CNVs, which are harder to measure, the standards the field requires are weaker."

The jury is out on how much CNVs matter for common diseases. A study this year[8] profiled 3,423 CNVs, or perhaps half of all those larger than 500 base pairs. It found that most not only don't explain much disease, but are also so closely associated with common SNPs that they've already been explored, albeit indirectly.

Scherer is not so sure. He was part of a team that resequenced a human genome and compared it to a reference. It found that the genome differed from the reference in only 0.1% of SNPs, but in 1.2% of CNVs. The

analysis indicated that up to one-quarter of CNVs are not associated with SNPs, and so are likely to be missed by SNP studies[9].

As with SNPs, larger effects may be found in rarer and harder-to-measure variants. Scherer has done studies showing that people with autism-spectrum disorders carry more rare CNVs than do controls. To be certain that the CNVs were correctly typed, he and his colleagues ran subsets of samples through calling algorithms that convert an instrument's signals into a sequence of base pairs, and used two platforms (by Illumina, of San Diego, California, and Agilent) to identify them[10].

Scherer says that many research groups are still learning about CNVs and don't fully realize the need to validate their data. "People are looking for low-hanging fruit; they see what they want to see and publish it," he says. The situation is improving, with the maturation of databases that collect diverse data on variation. "Now that we have much better data sets to compare to, it's becoming more accurate." **M.B.**

Seeing more SNPs

As genome-wide association studies (GWAS) get larger, the technical challenges pile up, and an onslaught of dense microarrays is compounding the issue by encouraging researchers to combine data sets. Genotyping a few-dozen single nucleotide polymorphisms (SNPs) in a sample is not much cheaper than genotyping hundreds of thousands, says Peter Donnelly, director of the Wellcome Trust Centre for Human Genetics in Oxford, UK. So rather than designing a targeted follow-up study on a handful of SNPs, researchers are more likely to try to replicate an association through meta-analysis, using samples that have been fully genotyped elsewhere. "That needs care," says Donnelly. Even in straightforward GWAS, everything that looks like a signal is probably an artefact, he says. Combining results typed on one platform in one lab and on another in

a different lab creates more opportunities for artefacts.

Even when cases and controls are processed by the same group, all the cases can be on one set of microarray plates and all the controls on another. This introduces potential for systemic error that sometimes leads to up to 30% of the data being discarded, says Christophe Lambert, chief executive of Golden Helix in Bozeman, Montana, which provides software and analytical services for genetic research. "Everyone is running these experiments and asking the statisticians to fix the problems, when a simple block randomization at the beginning could have fixed it."

Some problems occur before the sample is collected, says James Clough of Oxford Gene Technology, a genotyping-services firm. "Samples will be collected in multiple

centres and multiple countries." That can pose challenges when clinical standards vary. The best studies put more effort into collecting phenotypes than collecting samples, he says.

Careful characterization of phenotype could make genetic signals more apparent, says Greg Gibson, director of the Center for Integrative Genomics at the Georgia Institute of Technology in Atlanta. Many aspects of phenotype are extremely variable, so longitudinal measurements of factors such as blood-lipid levels, body-mass index or toxin exposure could control for transient effects and effectively boost genetic signals. GWAS could be more successful at implicating genes if they concentrate on qualities more closely tied to genetics, such as lipid levels or endophenotypes, he says. "Just mapping genotype to disease is several steps away from gene expression." **M.B.**

---

what it turns out to be. For example, even if all the genetic components of a disease were based in very many common variants with small effects, it would be good to know that." And even if the effects of variants of a gene in a general population are small, those of modulating that gene with a drug can be large. For instance, variations in the gene encoding 3-hydroxy-3-methylglutaryl coenzyme A reductase have been connected in GWAS with small effects[6] on cholesterol levels, but the statin drugs that modulate that gene product are effective and very widely prescribed. Although statins were not inspired by GWAS, such studies have turned up surprising connections with therapeutic implications, such as the role of the immune system in age-related macular degeneration, or of cell-cycle regulators in type 2 diabetes. In fact, says Altshuler, such results could be useful for focusing sequencing studies. "The genome-wide association paradigm might be that you find the gene using GWAS, and then sequence to find the rarer variants."

One of the biggest GWAS so far assessed samples from more than 100,000 individuals for more than 2 million SNPs, and identified 95 loci associated with variation in cholesterol and triglyceride levels in blood, 59 of which had never been reported before, and many of which were not near genes known to be associated with lipid metabolism[5]. Follow-up experiments in mice not only showed that some newly implicated genes had direct effects on plasmid lipid levels, but also identified a new cell-signalling pathway that could be targeted for therapeutic intervention. Another study[7] examined four genes that had been implicated by GWAS as contributing to high blood-triglyceride levels. Common variants explained less than 10% of

observed variation, so researchers sequenced the genes to identify rare missense and non-sense variants — two categories of mutations likely to change protein function. Nearly twice as many of these were found in affected individuals than in controls.

DIFFERENT STRATEGIES

The debate over the best approach for finding causal variants, says Altshuler, reflects researchers' various options for studying disease, and their limited funds. The decision whether to sequence a handful of samples or genotype thousands depends on whether researchers believe that a disease will be explained by a few rare variants or many common ones.

The answer will vary by disease. Current GWAS, for example, explain more heritability for autoimmune disorders and late-onset diseases such as Alzheimer's and heart disease than for mental conditions such as schizophrenia and autism. Natural selection suggests ready explanations, although they are hard to prove. Almost by definition, late-onset diseases tend to affect individuals in their post-reproductive years, and so are less likely to be selected against. And some genetic variants that contribute to one disease might actually be protective against others, and so could be favoured by natural selection. Genetic variants for sickle-cell anaemia, for example, can help to prevent carriers from contracting malaria, and there are hints that genes causing predisposition to some autoimmune diseases also confer resistance to infection.

In an effort to gather concrete evidence on which technologies are best suited to explaining the inheritance of common diseases, Altshuler has begun a study, with Mike Boehnke at the University of Michigan and Mark McCarthy at

the University of Oxford, to compare the same population using several techniques. In this case, the study will compare what Altshuler calls "extremes of risk": subjects who are at high risk for diabetes because of their age and weight but do not have the disease will be compared with slimmer, younger subjects who have been diagnosed with it. Presumably, individuals in the first group will carry relatively more protective variants, whereas those in the latter will have more susceptibility variants. About 2,600 people will be genotyped for 5 million SNPs, and be submitted to whole-exome and whole-genome sequencing.

Altshuler says that the study should not only uncover important information about diabetes, but also offer empirical data to help researchers choose the most appropriate technology, or combination of technologies. "We want to know what each approach finds that the others don't," he says. "Right now, no one actually knows which one is going to apply to which disease. Investigators have to take different bets." ∎

Monya Baker *is technology editor for* Nature *and* Nature Methods.

1. Park, J. H. *et al. Nature Genet.* **42,** 570–575 (2010).
2. International HapMap 3 Consortium *Nature* **467,** 52–58 (2010).
3. Thorleifsson, G. *et al. Nature Genet.* **42,** 906–909 (2010).
4. Ng, S. B. *et al. Nature Genet.* **42,** 30–35 (2010).
5. Teslovich, T. M. *et al. Nature* **466,** 707–713 (2010).
6. Burkhardt, R. *et al. Arterioscler. Thromb. Vasc. Biol.* **28,** 2078–2084 (2008).
7. Johansen, C. T. *et al. Nature Genet.* **42,** 684–687 (2010).
8. Wellcome Trust Case Control Consortium *Nature* **464,** 713–720 (2010).
9. Pang, A. W. *et al. Genome Biol.* **11,** R52 (2010).
10. Pinto, D. *et al. Nature* **466,** 368–372 (2010).

GENOTYPING, SEQUENCING AND/OR ANALYSIS

Company	Products/Activity	Location	URL	
454 Life Sciences	Next-generation sequencers	Bradford, Connecticut	www.454.com	
ACGT	SNP genotyping and sequencing services	Wheeling, Illinois	www.acgtinc.com	
Affymetrix	High-density customizable microarrays for genotyping	Santa Clara, California	www.affymetrix.com	
Agilent Technologies	Customizable, high-density microarrays for CNV genotyping	Santa Clara, California	www.chem.agilent.com	
Ambry Genetics	Next-generation sequencing and microarray services	Aliso Viejo, California	www.ambrygen.com	
Applied Biosystems	Next-generation sequencing	Carlsbad, California	www.appliedbiosystems.com	
ATLAS BioLabs	Microarray-based genomic services, targeted sequence capture and next-generation sequencing	Berlin, Germany	www.atlas-biolabs.de	
Beckman Coulter	High-throughput SNP discovery and resequencing	Brea, California	www.beckmancoulter.com	●
Complete Genomics	Large-scale genome-sequencing services	Mountain View, California	www.completegenomics.com	
CynerGene	SNP genotyping and fine-mapping services	Frederick, Maryland	www.cynergene.com	
deCODE genetics	Gene-discovery services, genotyping and sequencing facilities	Reykjavik, Iceland	www.decode.com	
Eureka Genomics	Next-generation sequencing and services, algorithms and data collections, resequencing and mapping services	Houston, Texas	www.eurekagenomics.com	
Expression Analysis	Genotyping assays, DNA sequencing services, sequence enrichment technologies and bioinformatics support	Durham, North Carolina	www.expressionanalysis.com	●
febit	Sequence capture for targeted resequencing, barcoding	Heidelberg, Germany	www.febit.com	
Fluidigm	PCR and SNP analysis on tiny chips	South San Francisco, California	www.fluidigm.com	
GeneDx	Genetic-testing services, mutation-confirmation services	Gaithersburg, Maryland	www.genedx.com	
Genizon Biosciences	High-throughput genotyping for CNVs and SNPs	St Laurent, Canada	www.genizon.com	
Genzyme Genetics	Human genetic-testing services	Cambridge, Massachusetts	www.genzymegenetics.com	
Helicos Biosciences	Targeted and whole-genome resequencing	Cambridge, Massachusetts	www.helicosbio.com	
iGenix	Screening services for SNPs and CNVs in customized assays	Bainbridge Island, Washington	www.igenixinc.com	
Illumina	Next-generation sequencing machines and services	San Diego, California	www.illumina.com	●
KBioscience	Whole-genome amplification and SNP genotyping	Hoddesdon, UK	www.kbioscience.co.uk	
Marligen Biosciences	Customized SNP genotyping services	Rockville, Maryland	www.marligen.com	
NABsys	Whole-genome sequencing technology	Providence, Rhode Island	www.nabsys.com	
National Center for Genome Resources	High-throughput SNP genotyping, sequencing and analysis	Santa Fe, New Mexico	www.ncgr.org	
Oxford Gene Technology	Genomic services, high-throughput microarray services	Oxford, UK	www.ogt.co.uk	
Oxford Nanopore Technologies	Sequencing platform based on nanopore technology	Oxford, UK	www.nanoporetech.com	
Pacific Biosciences	Single-molecule SMRT sequencing systems	Menlo Park, California	www.pacificbiosciences.com	
Polonator	Second-generation sequencing technologies	Salem, New Hampshire	www.polonator.org	
Precision Biomarker Resources	Automated, high-throughput microarray services	Evanston, Illinois	www.precisionbiomarker.com	
Raindance Technologies	Platform creates microdroplets for targeted sequencing	Lexington, Massachusetts	www.raindancetechnologies.com	
Roche NimbleGen	Arrays for genome capture	Madison, Wisconsin	www.nimblegen.com	●
Sequenom	SNP validation and fine-mapping studies	San Diego, California	www.sequenom.com	
Sequetech	DNA sequencing services	Mountain View, California	www.sequetech.com	
SeqWright	Next-generation sequencing, SNP resequencing and genotyping	Houston, Texas	www.seqwright.com	
Source BioScience Life Sciences	Genotyping and sequencing	Nottingham, United Kingdom	www.lifesciences.sourcebioscience.com	
US Genomics	High-grade DNA from complex samples, tagging, microfluidics and genomic mapping	Woburn, Massachusetts	www.usgenomics.com	

SOFTWARE AND ANALYSIS

Company	Products/Activity	Location	URL
Accelrys	Workflows for data management, analysis and reporting	San Diego, California	accelrys.com
BC Platforms	Data-management systems for genotyping and phenotyping	Espoo, Finland	www.biocomputing.com
BioDiscovery	Software for processing CNV and microarray data	El Segundo, California	www.biodiscovery.com
CLC Bio	Software developer for next-generation sequencing analysis	Aarhus, Denmark	www.clcbio.com
DNAnexus	Cloud-based informatics for next-generation sequencing data	Palo Alto, California	dnanexus.com
GATC Biotech	Next-generation sequencing services, DNASTAR software	Konstanz, Germany	www.gatc-biotech.com

SOFTWARE AND ANALYSIS

Company	Products/Activity	Location	URL
Gene Codes	Bioinformatics software for sequence analysis	Ann Arbor, Michigan	www.genecodes.com
Genedata	Bioinformatics systems for sequence and genome analysis, functional genomics	Basel, Switzerland	www.genedata.com
GenoLogics	Whole-genome SNP analysis and CNV analysis	Victoria, Canada	www.genologics.com
Genomatix	Algorithms and comprehensive databases for genomic data	Munich, Germany	www.genomatix.de
GenomeQuest	Informatics for next-generation sequencing data	Westborough, Massachusetts	www.genomequest.com
Geospiza	Data analysis of microarrays and next-generation sequencing	Seattle, Washington	www.geospiza.com
Golden Helix	Software and analytical services for SNP and CNV studies	Bozeman, Montana	www.goldenhelix.com
Hitachi Solutions	Services including sequencing and genotyping analysis	Tokyo, Japan	www.hitachi-solutions.com
InfoQuant	High-throughput CNV analytics and data management	London, UK	www.infoquant.com
Integromics	Data management and analysis for next-generation sequencing	Madrid, Spain	www.integromics.com
JMP Genomics	Software for various applications, including CNV	Cary, North Carolina	www.jmp.com
Laragen	DNA sequencing and genotyping software	Los Angeles, California	www.laragen.com
MiraiBio	Genotype analysis software and other genomics products	South San Francisco, California	www.miraibio.com
Nucleics	Software, reagents and services for DNA sequencing	Bendigo, Australia	www.nucleics.com
Ocimum Biosolutions	Services including SNP genotyping and CNV analysis	Hyderabad, India	www.ocimumbio.com
Paracel	Software for BLAST searches and genomic comparison	Pasadena, California	www.paracel.com
Partek	Microarray data analysis; software for next-generation sequencing	St Louis, Missouri	www.partek.com
Premier Biosoft	Software for various applications, including CNV	Palo Alto, California	www.premierbiosoft.com
Progeny Software	Genotype management for genome-wide association studies	Delray Beach, Florida	www.progenygenetics.com
Sage Bionetworks	Repository for data sets in integrative genomics	Seattle, Washington	www.sagebase.org
SAS	Statistical analysis of genetic-marker data	Cary, North Carolina	www.sas.com
Soft Genetics	Software tools for genetic analysis	State College, Pennsylvania	www.softgenetics.com

GENERAL

Company	Products/Activity	Location	URL	
ABgene	High-throughput PCR plates, reagents for nucleic-acid sequencing	Epsom, UK	www.abgene.com	
Alpha Laboratories	Reagents, plasticware and laboratory supplies	Eastleigh, UK	www.alphalabs.co.uk	
Bioneer	Nucleic-acid amplification, AccuPrep extraction and purification kits, enzymes	Alameda, California	bioneer.com	
Bio-Rad	Microarrays, nucleic-acid preparation, purification and amplification	Hercules, California	www.bio-rad.com	●
Biosearch Technologies	Custom oligonucleotide synthesis	Novato, California	www.biosearchtech.com	
Epicentre Biotechnologies	Sequencer-ready libraries, gene expression, enzymes	Madison, Wisconsin	www.epibio.com	
GenScript	PCR reagents	Piscataway, New Jersey	www.genscript.com	
Integrated DNA Technologies	PCR assay design tools, custom oligos, other reagents	Coralville, Iowa	www.idtdna.com	
PerkinElmer	SNP scoring and detection products	Waltham, Massachusetts	www.perkinelmer.com	
Promega	PCR products, purification products, polymerases and reagents	Madison, Wisconsin	www.promega.com	
Qiagen	Sample enrichment for next-generation sequences; PCR kits and reagents, PCR-based genotyping	Hilden, Germany	www.qiagen.com	
Rubicon Genomics	Pre-analytical amplification technologies for qPCR and next-generation sequencing, single-cell techniques	Ann Arbor, Michigan	www.rubicongenomics.com	
SA Biosciences	Beadchips that detect SNPs from a variety of samples	Frederick, Maryland	www.sabiosciences.com	
Sigma-Aldrich	Reagents and products for PCR and arrays	St Louis, Missouri	www.sigmaaldrich.com	●
Takara Bio	Products for PCR, including enzymes and thermocyclers	Shiga, Japan	www.takara-bio.com	
Thermo Fisher Scientific	Equipment, enzymes and reagents	Waltham, Massachusetts	www.thermofisher.com	
Transgenomic	High-sensitivity genetic variation and mutation analysis using PCR	Omaha, Nebraska	www.transgenomic.com	

● see advertisement

CAREERS

D. AUSSERHOFER

The Philological Library at the Free University of Berlin was built in 2005 as part of a continuing renovation of the campus.

REGIONS

Science city chic

Berlin is an international hotspot for young scientists. Now it has to provide the incentives to help them stay long term.

BY QUIRIN SCHIERMEIER

Francesca Spagnoli, a native of Italy, spent seven years as a postdoc at Rockefeller University in New York. Two years ago, looking for an opportunity to start her own research group, she moved to Berlin, where she has since secured a group-leader position at the Max Delbrück Centre for Molecular Medicine (MDC). In a sense, the 40-year-old stem-cell researcher is a typical Berlin scientist. Although she enjoys the city's charms, and benefits from its science strengths, she is aware that the scarcity of permanent positions here will sooner or later force her to move on. "Berlin is definitely one of the best places for science in Europe now," she says. "But as tenure is rare here, I will probably have to move again."

Berlin's international flair and relatively cheap living — unlike in London, Munich or Paris, a decent two-room apartment can be leased for about €400 (US$550) per month — appeals to artists, hipsters and scientists alike. Young scientists come in search of career springboards at the numerous labs in and around the capital. The sheer density of science in the region is impressive: Berlin and nearby Potsdam host four large research universities, eight Max Planck institutes, three national research centres, a well-established biotechnology industry and the headquarters of several internationally operating pharmaceutical companies.

"In terms of potential and creativity, Berlin need not fear comparison with emerging science cities such as Singapore or Shanghai," says Detlev Ganten, founding director of the MDC. The region's strong points, he says, are its strengths across many fields, from innovative research in astrophysics and cosmology at the Albert Einstein Institute in Potsdam to strong polymer and materials research centres, and the MDC's research on cells, tissues, organisms and individual diseases.

In general, scientists in Berlin benefit from the country's good funding opportunities. Germany's science budget has grown faster than those of most other countries in and outside Europe (see *Nature* **467**, 499–500; 2010), and it is complemented by European funding. Just one year into her job at the MDC, for example, Spagnoli was awarded one of the prestigious — and, at €1.65 million, generous — starting grants by the European Research Council (ERC).

Yet, despite the city's science assets, its continued emergence as a science powerhouse is not a given. Berlin hosts no research institution in the same league as the top-ranked academic institutions in Britain and the United States. Ranked 178 in the *Times Higher Education* (*THE*) magazine's annual list of the world's top 200 universities in October, Humboldt University lags far behind the Harvards, Cambridges ▶

and Oxfords of the global science community. So does the Free University of Berlin — one of nine German 'elite' universities, which each receive €50 million a year in extra support from the federal science ministry's 'excellence initiative'. It did not even make the *THE* list.

And there are few tenure-track positions, meaning little long-term security for researchers. Until recently, Germany had no tenure-track system at all. It has become an option in the past decade, but remains rare. Advancement in German universities often comes by a complicated procedure that lacks transparency.

Meanwhile, public debts and budget constraints continue to plague Berlin. Perhaps the biggest challenge is the precarious institutional funding situation in medical science. On the basis of the government's own funding standards for medical science, Berlin has not fared well. The Charité, a multi-campus medical school for both Humboldt and the Free University, had to cut almost €250 million from its budget for 2005–10. This means that overdue investments in building and renovation had to be repeatedly postponed. It hasn't yet hampered the science much, says Ganten, who's also the former head of the Charité, but he fears it might in the long term. To make the most of its strong points and to continue to attract talent, Berlin will have to sustain financial support, overcome some political wrangling and create more tenure-track positions to convince more young scientists to stay in the city.

POLITICAL SUPPORT

Members of several political parties in Berlin's senate, including the Christian Democrats, the Greens and the Free Democrats, are keen to establish Berlin as a major force in the European science landscape. This year, the Berlin senate eventually approved €330 million for the Charité, but renovations of the rundown main Charité clinic in Berlin–Mitte alone would cost some €600 million. The main issue is whether the city can afford to maintain all three Charité sites. In June, Jürgen Zöllner, the Social Democrat science senator, said that all three campuses will remain open. But the number of patient beds in the main clinic in Berlin–Mitte will be cut by 500, starting in 2012.

Nonetheless, Berlin policy-makers and university administrators understand how important science is for its future development, says Ganten. "What's lacking here," he says, "is a smart one-stop science marketing scheme of the kind that our Asian competitors master so well." Ganten would like to see lasting financial support, including a targeted programme to attract high-profile foreign scientists to Berlin — something akin to the success at Singapore's Biopolis. One problem, he says, is that the Max Planck institutes, Berlin's universities and the Helmholtz centres rarely collaborate.

The Einstein Foundation, established in 2009, could help to remedy that lack of crosstalk. As a sort of umbrella organization for

Berlin science, it aims to support the state's research both financially and structurally. But the foundation is already troubled by political infighting. In July, Berlin's senate criticized its managers' high salaries. The foundation is to provide more than €40 million for selected science projects in Berlin, but it is not yet clear where the money will come from.

EARLY-CAREER ASPIRATIONS

Like Spagnoli, Berlin's many young scientists enjoy the capital's lifestyle and reasonable cost of living, and are so far unhampered by the political disputes in the state's science ministry over budgets and priorities. "Berlin makes it easy for newcomers," says Spagnoli. "Language is no barrier, my husband has found a nice job, and renting an apartment was no problem at all."

"Attracting foreign talent to Berlin has become easy," says Leif Schröder, a group leader at the Leibniz Institute for Molecular Pharmacology. Schröder, who this year also received a €1.5-million ERC starting grant, is developing magnetic resonance imaging techniques for biomarkers in different diseases. His new group comprises Australian, Italian and German postdocs and PhD students. "People started approaching me and suggesting research ideas and collaborations as soon as I arrived. It's pleasing to see what's happening here."

> *"Berlin need not fear comparison with emerging science cities such as Singapore or Shanghai."*
> Detlev Ganten

"Berlin hosts a huge pool of scientists from which to choose potential collaborators," says Ingrid Hotz, an expert on data analysis and visualization at the state-funded Zuse Institute Berlin, which provides advanced computing services for many scientific applications. Hotz leads an independent junior research group funded by the DFG, Germany's main grant-giving agency. She came to Berlin in 2006, after a three-year stint at the University of California, Davis, and maintains collaborations with groups in hydrodynamics, medicine, geology and gravitational physics at local institutions such as the Max Planck Institute for Gravitational Physics and the Charité. "I am turning other people's data and experiments into images, so teaming up is everything for me," she says. "Not all collaborations bear fruit, but fortunately there are more than enough potential research partners around here."

This is equally true in Potsdam, located less than an hour's train ride from Berlin Zoo. Potsdam complements the capital's science base. With around 6,000 academic scientists working at Potsdam's universities, along with the Max

Planck, Helmholtz and Fraunhofer institutes and the Potsdam Institute for Climate Impact Research, Potsdam has the highest density of researchers anywhere in Germany.

FUTURE GROWTH

Despite some funding problems, notably the case of the Charité, there are signs that the Berlin–Brandenburg research base will continue to grow in the next few years. The Berlin Institute for Medical Systems Biology, now on the MDC's Berlin–Buch campus, will relocate in 2015 to the Humboldt University's main campus in Berlin–Mitte. The move will substantially expand the MDC's — and Berlin's — systems-biology capacities. More than 20 new research groups totalling about 100 scientists, jointly funded by Berlin and the federal government, are to be recruited over the next few years.

The MDC will also become the hub of the planned German centre for cardiovascular research, a mostly federal-government-funded programme involving universities and research institutes across the country. The Berlin School of Public Health, a joint programme by Berlin's universities aimed at midcareer health professionals, mainly focuses on master's students with degrees in public health and epidemiology. Students often go on to leadership positions in government, research and non-profit agencies. Ganten hopes that the city can build a reputation as a global 'public-health capital'.

Growth is less certain within the Berlin–Brandenburg biotech cluster, which, with 82 biotech companies, is the largest in Germany. The sector is in a consolidation phase. In 2009, for example, the US laboratory giant Thermo Fisher acquired the diagnostics firm Brahms in Hennigsdorf near Berlin for €330 million.

Although the climate for new jobs in the biopharmaceutical field has cooled down, some biotechs are still offering attractive jobs, says Steffen Goletz, chief executive and founder of Glycotope, a biotech company based in Berlin and Heidelberg. Glycotope, which specializes in therapeutic antibodies and non-antibody proteins, has hired more than 100 researchers, engineers and technicians in the past few years, many from Berlin research institutes, says Goletz.

Spagnoli is mindful of such local opportunities, but has not decided what her next career step will be. "I will probably move, but who knows?" she says, adding that she hopes excelling in her science will open multiple doors. ∎

Quirin Schiermeier is Nature's Germany correspondent.

CORRECTION

In the 'By The Numbers' on Belgium (*Nature* **467,** 876; 2010), the Catholic University of Louvain was wrongly depicted as being in Flanders. It is actually in Wallonia.

TURNING POINT
Greg Graffin

TANKWART

Greg Graffin has two passions: evolutionary biology and music. The latter led him to co-found renowned punk band Bad Religion in 1980. But even as a child he was wooed by the provocative lyrics of Charles Darwin's theory on evolution. Now an occasional biology lecturer at the University of California, Los Angeles, Graffin last month released the book Anarchy Evolution: Faith, Science, and Bad Religion in a World without God, *co-authored with science writer Steve Olson. He calls the work part memoir and part polemic. Graffin tells* Nature *how he turned to music without ever abandoning science.*

When did you decide to pursue a PhD in evolutionary biology?
On entering college. I had already released an album, in 1982. I was only 17 or 18, and the band was my outlet, but it didn't yet have an international reputation. When I was a teenager, science meshed with my developing ideals — such as the challenge to authority that was central to punk rock. In science, anyone from any walk of life could make a discovery that would overturn prevailing hypotheses. And that was a cause for celebration among scientists. It taught me that challenging authority has good results.

How did you get a PhD while the lead singer for an internationally known rock band?
I got interested in palaeontology and vertebrate history — sparked by books on human evolution — then vertebrate evolution. Studying with palaeontologists kindled my interest in fieldwork.

I struggled to keep one foot in music and one in academia. I had worked on my PhD for three years full time before I realized Bad Religion could be a legitimate career. We had tour offers from 12 countries.

How did science influence your music?
One example might be our song *I Want Something More*. The lyrics discuss how we as humans struggle to form our world view. That is why religion was such an easy target — not to tear it down, but to identify its fatal flaws. It claims to offer a world view, but not one that resonates with us.

Did touring hinder your PhD?
I took six years' leave from graduate school, but it wasn't just because of the band. I had children, a divorce. All the while, I considered Bad Religion a fruitful intellectual pursuit. By the time I went back to graduate school, my focus had shifted from vertebrate palaeontology to the intersection of science and religion.

Was the band supportive of your science?
The band has been together for 30 years. Everyone has always known that touring follows the academic schedule. We tour in summer when there is no fieldwork. Since high school, we've recognized that what makes us unique is the stories under the music, and my science is among them.

Do scientists often not take you seriously?
Yes. I'm more likely to be criticized for my science, because I've been successful in music. I appreciate criticism, but so much of it isn't constructive. Steve Olson has given me insight into what to expect with the book. Academic scientists aren't generally interested in books for the public. So when one comes out, the authors can't expect much praise from scientists. My goal both as a singer and an instructor is to educate through provocation and entertainment.

How do your students react to your fame?
My classes have up to 350 students. The usual attitude is, "We're serious. We're pre-meds. Just tell us what we need to get an A." With pre-meds, you will weed out a lot of the punk rockers. But I have lively conversations in office hours.

Did you really, as noted in the book, forgo an outing with a groupie to do fieldwork in Brazil?
By the time that happened, I had already had the experiences that a rock star needs to experience. I started Bad Religion at 15. And those activities stimulate different parts of my brain. ∎

INTERVIEW BY GENE RUSSO

Grant system simplified

The science-research grant-reporting and audit requirements of the European Commission (EC) are to become simpler. New measures include accepting less-stringent grantee accounting practices, discarding rules requiring grant recipients to deposit their grant money into a bank account and improving transparency for criteria and timelines throughout the grant cycle. In a meeting on 11–12 October, the EC's Competitiveness Council requested that the commission streamline the process during the Seventh Framework Programme and make further changes for the Eighth Framework Programme, which launches in 2014. A spokesman says that the aim is to cut red tape and introduce a more flexible and user-friendly system without decreasing financial control and oversight.

Biodiversity centre bid

The German Research Foundation (DFG), the country's main grant agency, is soliciting proposals from universities to host a biodiversity centre that could create up to 80 postdoc and PhD positions, 6 professorships and 10 group-leader positions. The DFG will make its selection by April 2012; the host university will build a new centre or expand an existing lab. Proposals must be submitted by 14 January 2011, and applicants must show how they will collect and analyse data and set up a study programme. The DFG will provide between €4 million (US$5.5 million) and €7 million a year for up to 12 years. The successful applicant must have expertise in biodiversity theory and modelling, ecology, evolution and the science of conservation.

Gender target missed

European Union institutions and nations have not met benchmarks for women's participation in the research workforce, a report says. In an initiative to examine gender issues in science, the European Commission decided in 1999 that women should make up 40% of its panels, and in 2005 that 25% of senior researchers should be female. The initiative's final report, *Stocktaking 10 years of 'Women in Science'*, released on 13 October, identifies reasons for women's low numbers, including inconsistent political support. Progress on gender issues continues to be erased by changes to political leadership, it says.

THE GREATEST SCIENCE-FICTION STORY EVER WRITTEN

A real page-turner.

BY ERIC JAMES STONE

I tore open the self-addressed, stamped envelope and unfolded the single sheet of paper inside. The letter was signed by the editor of *Analog Science Fiction* and was addressed to me, personally, which still gave me a warm feeling after all those years of form rejections. But what I craved now was an acceptance.

And … this wasn't it. *Good luck placing this elsewhere*, the letter read.

I shoved the rejection in my overstuffed file with the rest of them. Eyeing the four-inch-thick wad of paper, I felt a wave of despair. Maybe I didn't have what it took to be a science-fiction writer. Maybe I should just give it up — after all, I worked for a quantum-computing start-up. That was almost science fiction, even if all I did was manage the website. Maybe that was as close as I'd ever get.

The next day, while having a mint Oreo shake at a restaurant near my office, I told Caleb, one of the quantum-circuit experts I worked with, that I doubted I'd ever see my name in print.

"Don't quit," he said. "You're a great writer." He'd read a few of my stories to give me feedback on where I'd got the science wrong.

I shrugged. "Doesn't matter, if I'm not writing what editors want to buy."

"Why don't you?"

"Why don't I? It's not that easy," I said. "There's no way of knowing what an editor will like. I write the best story I can, but apparently that's just not good enough."

"So it's subjective." Caleb took a bite of his burger and chewed thoughtfully.

"Yeah," I said, playing with the last spoonful of shake in my cup. "What one editor thinks isn't worth publishing, another might think is the greatest science-fiction story ever written. It's just my luck that the editor who would love my stuff isn't actually an editor anywhere."

"No, no," Caleb said. "You're looking at it all wrong. What you need is a story that adapts itself perfectly to the editor."

I dabbed my lips with a paper napkin. "I just told you I don't know how to write what they're looking for."

"Right." Caleb grabbed the napkin from my hand, flattened it out, took a pen from his pocket and sketched a curve. "It's a probability function. The right combination of words makes them buy the story, the wrong combination means they don't."

"I suppose," I said dubiously.

"And if it's a probability function, then our quantum computer can handle it." He scribbled an equation, crossed part of it out, then added something. "Oh, boy. This will revolutionize publishing."

I stared at him. "What are you talking about?"

He stopped scribbling. "Imagine you open a book, and from the very first word, it's exactly what you want to read. Every word is perfect, the characters fascinate you, the plot thrills you …"

"That'd be cool," I said.

"And someone else opens their copy of the same book, and it's perfect for them. Only if you compare the two books, the words aren't the same. The story and characters aren't even the same. The book has adapted itself to be the perfect book for whoever first opened it."

I frowned. "You mean, it's like an e-book that changes based on personal preferences?"

"No, this would be printed on paper. But the text itself would have been composed using a quantum computer, like the one we have at the office, using a program to create a quantum probability wave function that doesn't collapse until someone actually observes what was printed." Caleb sat back with a satisfied grin.

"And when the wave collapses …" I said, not quite sure that I understood the implications.

"The book becomes the best book ever written for whoever collapses the wave. It's brilliant." Caleb leaned forward. "And we can use it to make sure you get your name in print. How would you like to be the author of the greatest science-fiction story ever written?"

I stared at the sheets of paper lying facedown on the printer. "You're certain I can't take just a peek?"

"If you do," Caleb said, "the wave function will collapse and the story will become the best story for you, not for the editor of *Analog*. He needs to be the one to see it first."

"Can I at least know the title?" I felt kind of awkward submitting a story that I knew nothing about, even though Caleb assured me that I could still be considered the author, as the computer could not have been programmed to create a probability wave function for science-fiction stories without my help.

"Nope," he said. "I've hard-coded your name and contact information into the printout, but the rest remains undecided until the editor reads it."

With a sigh, I slid the manuscript into the manila envelope and sealed it.

Sixty days later, my SASE returned. I took it unopened to the office the next day — I wanted to open it with Caleb.

"Could be an acceptance or a rejection," I said.

"Open it," Caleb said, looking at the envelope. "You have to collapse the wave function. But I'm sure it's an acceptance."

I opened it.

"Read it out loud," Caleb said.

I looked past my name and began reading. "In my opinion, this is the greatest science-fiction story ever written." My heart leapt within me, and I continued. "It is undoubtedly the best story you have ever submitted to me. But what on Earth made you think you could get away with submitting a verbatim copy of 'Nightfall' by Isaac Asimov?" ∎

A Writers of the Future contest winner, **Eric James Stone** *has had stories published in* Analog, InterGalactic Medicine Show *and various other venues. His website is www.ericjamesstone.com.*

JACEY

Volatile accretion history of the Earth

It has long been thought that the Earth had a protracted and complex history of volatile accretion and loss[1,2]. Albarède[3] paints a different picture, proposing that the Earth first formed as a dry planet which, like the Moon, was devoid of volatile constituents. He suggests that the Earth's complement of volatile elements was only established later, by the addition of a small veneer of volatile-rich material at ~100 Myr (here and elsewhere, ages are relative to the origin of the Solar System). Here we argue that the Earth's mass balance of moderately volatile elements is inconsistent with Albarède's hypothesis but is well explained by the standard model of accretion from partially volatile-depleted material, accompanied by core formation.

Albarède follows standard practice by grouping volatile elements according to the temperatures at which they would be 50% condensed from a gas of solar composition, T_{50}. Particularly important to the discussion are Pb ($T_{50} = 727$ K) and other elements that condense in the temperature interval 500–800 K (Fig. 1). Lead is special because its isotopic composition reflects the time-integrated U/Pb ratio, owing to the decay of ^{238}U and ^{235}U to isotopes of Pb. In the silicate Earth, the Pb isotopic composition provides evidence for a major episode of U/Pb fractionation at 50–150 Myr (ref. 4).

Albarède argues that this young U–Pb age for the Earth marks the arrival of a 'late veneer' of volatile-rich material, which provided >99% of terrestrial Pb, at 50–150 Myr. To set the age, this requires that some of the late accreting Pb was volatilized during collisions, thereby decoupling the relationship between U/Pb ratio and Pb isotopic composition. Before this, during the main phase of accretion and core formation that Albarède places before 30 Myr, the Earth would have had an extremely high U/Pb ratio because it formed from material that was essentially devoid of elements with volatilities similar to or higher than Pb.

The standard alternative to Albarède's model[4] is based on terrestrial accretion from volatile-depleted material combined with partitioning of siderophile elements into the core. Lead exhibited siderophile

character and was partially lost to the segregating metal while the lithophile U was completely retained in the mantle. The loss of Pb to the core produces a silicate Earth with high U/Pb and a Pb isotopic composition that provides an average age for accretion and core formation. The young U–Pb age of the Earth relies on the segregation of some core material during the Moon-forming giant impact[5], which occurred at >60 Myr (ref. 6). The impact would set the age of U/Pb fractionation, provided that the added metal experienced at least partial equilibration with the silicate mantle.

Our principal objection to Albarède's hypothesis relates to the inconsistent mass balance of the model. In particular, the time of U/Pb fractionation (50–150 Myr) cannot date addition of the late veneer because this requires an implausibly large amount of material. The concentration of ^{204}Pb in CI chondrites (the most volatile rich meteorites) is 42 p.p.b. (ref. 7) and in the silicate Earth it is 2.5 p.p.b. (ref. 7; Fig. 1). Supplying the latter as a late veneer of CI chondrite composition would add 6% to the mass of the primitive mantle and deliver Re, Au, S, C and water at levels that are 5, 7, 11, 13 and 5–20 times greater, respectively, than found in the silicate Earth. It would also generate an Earth–Moon difference in W isotopic composition of 0.3ε and a difference in δ^{18}O of 0.5‰, significantly different from observations[6,8].

This inconsistency is increased if the late veneer was more volatile-depleted than CI chondrites or if Pb was partially lost from this material by volatilization. For example, Re–Os systematics indicate that the late veneer may have had a composition similar to that of H chondrites[9]. In this case, the ^{204}Pb mass balance requires addition of ~60% of the mass of the Earth's mantle and this would deliver Re, Au, S and C at respective levels that are 90, 100, 40 and 4 times those found in the silicate Earth. It would also generate a chondritic W isotopic composition (-1.9ε) and an Earth–Moon difference in Δ^{17}O of 0.28‰, both inconsistent with published results[8,10].

Our secondary comment is that both the Allende carbonaceous chondrite and the silicate Earth are substantially depleted in volatile elements relative to CI chondrites (Fig. 1). The Earth differs from Allende, however, because it is most strongly depleted in those volatile elements which are also siderophile (for example, S, Se, Au, Ge, Bi) and would hence have partitioned into the core. Allende, which has not experienced core formation, shows no depletion of these elements relative to lithophile elements of similar volatility. The most plausible explanation for this difference is that volatile elements were added to Earth during the principal phase of accretion and core segregation, and not only as part of a late veneer.

We conclude, therefore, that the mantle abundances of most moderately volatile elements, including Pb, were strongly affected by core segregation and only marginally altered by contributions from a late veneer. This does not imply that a late veneer (which enhanced the budgets of some highly volatile, and siderophile, elements) did not accrete. The U–Pb age of the Earth, however, was essentially unaffected by any such accretion and largely reflects the earlier process of core formation.
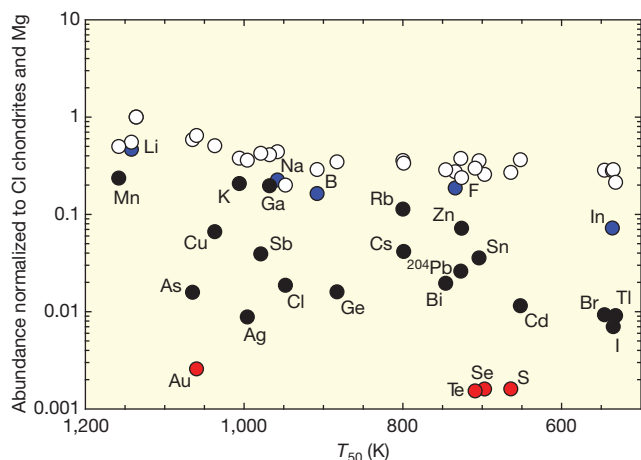
Figure 1 | Element abundances in the silicate Earth and the Allende meteorite. Shown are abundances of elements in the Allende CV3 chondrite (open circles)[11] and those in the silicate Earth[7] (filled circles) relative to CI chondrites (with values scaled to Mg/Mg$_{CI}$ = 1.0), plotted as a function of T_{50} (from ref. 12). Red, black and blue symbols represent respectively elements that are highly siderophilic, moderately to weakly siderophilic and (as far as is known) lithophilic.

B. J. Wood[1], A. N. Halliday[1] & M. Rehkämper[2]
[1]Department of Earth Sciences, Oxford University, Parks Road, Oxford OX1 3PR, UK.
[2]Department of Earth Science & Engineering, Imperial College, London SW7 2AZ, UK.
e-mail: berniew@earth.ox.ac.uk

BRIEF COMMUNICATIONS ARISING

1. Ringwood, A. E. Chemical evolution of the terrestrial planets. *Geochim. Cosmochim. Acta* **30,** 41–104 (1966).
2. Wänke, H., Constitution of terrestrial planets. *Phil. Trans. R. Soc. Lond. A* **303,** 287–302 (1981).
3. Albarède, F. Volatile accretion history of the terrestrial planets and dynamic implications. *Nature* **461,** 1227–1233 (2009).
4. Allègre, C. J., Manhes, G. & Gopel, C. The age of the Earth. *Geochim. Cosmochim. Acta* **59,** 1445–1456 (1995).
5. Wood, B. J. & Halliday, A. N. Cooling of the Earth and core formation after the giant impact. *Nature* **437,** 1345–1348 (2005).
6. Touboul, M., Kleine, T., Bourdon, B., Palme, H. & Wieler, R. Late formation and prolonged differentiation of the Moon inferred from W isotopes in lunar metals. *Nature* **450,** 1206–1209 (2007).
7. Palme, H. & O'Neill, H. S. C. in *The Mantle and Core* Vol. 2 (ed Carlson, R. W.) 1–38 (Elsevier, 2003).
8. Wiechert, U. *et al.* Oxygen isotopes and the Moon-forming giant impact. *Science* **294,** 345–348 (2001).
9. Walker, R. J. *et al.* Comparative ^{187}Re-^{187}Os systematics of chondrites: implications regarding early solar system processes. *Geochim. Cosmochim. Acta* **66,** 4187–4201 (2002).
10. Yin, Q. Z. *et al.* A short timescale for terrestrial planet formation from Hf–W chronometry of meteorites. *Nature* **418,** 949–952 (2002).
11. Wasson, J. T. & Kallemeyn, G. W. Compositions of chondrites. *Phil. Trans. R. Soc. Lond. A* **325,** 535–544 (1988).
12. Lodders, K. Solar system abundances and condensation temperatures of the elements. *Astrophys. J.* **591,** 1220–1247 (2003).

LETTER

Quantum tunnelling of the magnetization in a monolayer of oriented single–molecule magnets

M. Mannini[1,2], F. Pineider[1], C. Danieli[3], F. Totti[1], L. Sorace[1], Ph. Sainctavit[4], M.-A. Arrio[4], E. Otero[5], L. Joly[6]†, J. C. Cezar[7], A. Cornia[3] & R. Sessoli[1]

A fundamental step towards atomic- or molecular-scale spintronic devices has recently been made by demonstrating that the spin of an individual atom deposited on a surface[1], or of a small paramagnetic molecule embedded in a nanojunction[2], can be externally controlled. An appealing next step is the extension of such a capability to the field of information storage, by taking advantage of the magnetic bi-stability and rich quantum behaviour of single-molecule magnets[3–6] (SMMs). Recently, a proof of concept that the magnetic memory effect is retained when SMMs are chemically anchored to a metallic surface[7] was provided. However, control of the nanoscale organization of these complex systems is required for SMMs to be integrated into molecular spintronic devices[8,9]. Here we show that a preferential orientation of Fe_4 complexes on a gold surface can be achieved by chemical tailoring. As a result, the most striking quantum feature of SMMs—their stepped hysteresis loop, which results from resonant quantum tunnelling of the magnetization[5,6]—can be clearly detected using synchrotron-based spectroscopic techniques. With the aid of multiple theoretical approaches, we relate the angular dependence of the quantum tunnelling resonances to the adsorption geometry, and demonstrate that molecules predominantly lie with their easy axes close to the surface normal. Our findings prove that the quantum spin dynamics can be observed in SMMs chemically grafted to surfaces, and offer a tool to reveal the organization of matter at the nanoscale.

The canonical features of molecular clusters behaving as SMMs are a ground state with a giant spin and an easy-axis magnetic anisotropy[4]. For this reason, many SMMs so far reported contain highly anisotropic metal ions, such as Mn(III). Although single-molecule experiments based on nanojunctions have been described for the archetypal SMM, Mn_{12} (refs 10, 11), the intrinsic fragility of these polynuclear coordination compounds has prevented a significant advance in their organization on surfaces[12,13]. These severe drawbacks have been over-come in a class of tetranuclear Fe(III) clusters, Fe_4, which have a pro-peller shape (Fig. 1a)[14]. A major advantage of Fe_4 clusters is in the possibility to reinforce and functionalize their molecular structure using tripodal ligands, allowing thermal evaporation processing[15] as well as the preparation of single-molecule devices[16].

A ligand made up of a long aliphatic chain (hereafter C_9; Fig. 1b) and terminated by a sulphur-containing moiety has been used to graft Fe_4 molecules (in a compound hereafter denoted Fe_4C_9) to a Au(111) sur-face, allowing the demonstration by X-ray magnetic circular dichroism (XMCD) of the persistence of hysteresis in the magnetization[7]. A deeper surface characterization has revealed that the sulphur atoms of either one or both ligands can bind the surface (Fig. 1c), resulting in a random orientation of the easy magnetization axes of the grafted molecules[17]. A linker with only five carbon atoms in the chain (Fig. 1b) is expected to preclude grafting to the gold surface through both ligands, owing to the steric hindrance of the Fe_4 discoid (Fig. 1c).

The new derivative Fe_4C_5, whose complete formula is $[Fe_4(L)_2(DPM)_6]$ (where H_3L is 7-(acetylthio)-2,2-bis(hydroxymethyl)heptan-1-ol and HDPM is dipivaloylmethane), has therefore been prepared in pure, crys-talline form and chemically, structurally and magnetically characterized (Supplementary Methods, Supplementary Table 1 and Supplemen-tary Figs 1–6). The magnetic properties of the crystalline phase are typical for this class of compounds and are only briefly described here. The antiferromagnetic exchange interaction, J, between the central and the peripheral high-spin Fe(III) centres ($S_i = 5/2$) has been evaluated from the temperature dependence of the magnetization (Supplemen-tary Fig. 2) and found to be $J/k_B = 24.09(6)$ K, where k_B is the Boltzmann constant and the exchange Hamiltonian between sites i and j is written as $\mathcal{H}_{ij} = J\hat{\mathbf{S}}_i \cdot \hat{\mathbf{S}}_j$ ($\hat{\mathbf{S}}_i$ and $\hat{\mathbf{S}}_j$ are spin vector operators). As a consequence, the spin ground state has $S = 5$ and the first excited states are two degenerate $S = 4$ manifolds which lie $\sim$60 K above the ground state. A moderate Ising-type magnetic anisotropy lifts the degeneracy of the $S = 5$ state in zero applied field, as demonstrated by the low-temperature field dependence of the magnetization of a microcrystalline sample (Supplementary Fig. 3) and by high-frequency electron paramagnetic resonance spectra of a frozen
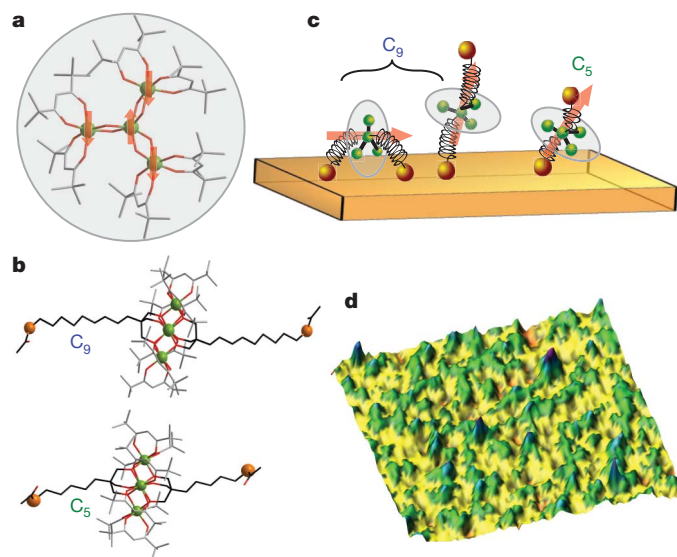
Figure 1 | Structure of Fe_4 clusters. a, Core structure of Fe_4 clusters with the estimated steric hindrance of DPM groups (grey disk) and the ferrimagnetic spin arrangement arising from Fe(III) ions. **b,** Complete molecular structure of the Fe_4C_9 and Fe_4C_5 systems as determined by single-crystal X-ray diffraction. **c,** Sketch of the grafting modes expected for Fe_4C_9 and Fe_4C_5; easy magnetization axes are represented by red arrows. **d,** Three-dimensional representation of a 50 nm × 50 nm scanning tunnelling microscope image with molecular resolution.

[1]Department of Chemistry 'Ugo Schiff' and INSTM Research Unit, University of Florence, 50019 Sesto Fiorentino, Italy. [2]ISTM-CNR, URT Firenze, University of Florence, 50019 Sesto Fiorentino, Italy. [3]Department of Chemistry and INSTM Research Unit, University of Modena and Reggio Emilia, 41100 Modena, Italy. [4]Institut de Minéralogie et de Physique des Milieux Condensés, CNRS UMR7590, Université Pierre et Marie Curie, 75252 Paris, France. [5]Synchrotron Soleil, Saint Aubin, 91192 Gif sur Yvette, France. [6]Swiss Light Source, Paul Scherrer Institut, 5232 Villigen, Switzerland. [7]European Synchrotron Radiation Facility, 38043 Grenoble, France. †Present address: IPCMS – DMONS, 67034 Strasbourg, France.

solution (Supplementary Fig. 6). The leading, second-order terms in the spin Hamiltonian, which describe magnetic anisotropy, are $D/k_B = -0.60(1)$ K and $E/k_B = \pm 0.020(5)$ K:

$$\mathcal{H}_{an} = D\hat{S}_z^2 + E\frac{1}{2}(\hat{S}_+^2 + \hat{S}_-^2)$$

where $\hat{S}_z$ and $\hat{S}_\pm$ are the spin z component and ladder spin operators, respectively. The observation of a small, but non-zero, E term indicates that the Fe_4C_5 molecules do not have rigorous three-fold symmetry. The SMM behaviour has been confirmed by measuring the alternating-current susceptibility in zero static field (Supplementary Fig. 4). Application of the Debye model shows that at temperatures as low as 1.7 K the relaxation of the magnetization follows the Arrhenius law, $\tau = \tau_0 \exp(U_{eff}/k_B T)$, with $\tau_0 = 0.061(2)\,\mu s$ and $U_{eff}/k_B = 14.8(1)$ K (Supplementary Fig. 5), and is thus in good agreement with the energy spreading of the $S = 5$ multiplet in zero field, $(|D|/k_B)S^2 \approx 15$ K.

We prepared the monolayer of Fe_4C_5 on gold by incubating a freshly annealed gold substrate for 20 h in a 2 mM solution of Fe_4C_5 in dichloromethane. Then we rinsed the sample several times with pure dichloromethane to remove any physisorbed material and finally dried it under inert atmosphere. A morphologic investigation using scanning probe microscopy at room temperature ($20 \pm 2\,^\circ$C) confirmed surface decoration by isolated molecules (Fig. 1d; see Supplementary Methods and Supplementary Fig. 7).

We investigated the electronic and magnetic properties of the monolayer by X-ray spectroscopy in the total-electron-yield detection mode[18], to achieve the surface sensitivity required to probe a monolayer of molecules[7,19]. In Fig. 2a, we show the X-ray absorption spectra (XAS) at the iron $L_{2,3}$ edges recorded at $T = 650(50)$ mK and in a magnetic field of 30 kOe with opposite circular polarizations of X-rays (left-polarized (σ^+) and right-polarized (σ^-)), along with the dichroic magnetic component, which is defined as the difference $\sigma^- - \sigma^+$. We note that XMCD probes the magnetic polarization of the material without being sensitive to the volume density of magnetic flux, in contrast with traditional magnetometry based on induction methods. This makes XMCD a unique tool for investigating magnetic molecules arranged in a bidimensional structure. The observed XAS and XMCD features are almost identical to those detected on the Fe_4C_9 monolayer[7], confirming the redox stability of the clusters on deposition. In particular, the L_3-edge XMCD region presents the expected fingerprint of ferrimagnetic spin arrangement, as supported by theoretical estimations[13].

To demonstrate any preferential orientation of the molecules on the surface, we investigated the X-ray natural linear dichroism[20] (XNLD) of the Fe_4C_9 and Fe_4C_5 monolayers. The XNLD signal is defined as the difference, $\sigma_V - \sigma_H$, between the cross-sections measured with vertically polarized (σ_V) photons and with horizontally polarized (σ_H) photons (Fig. 2b, inset). The experiment revealed well-defined spectral features for the Fe_4C_5 monolayer only (Fig. 2b), suggesting that in this case the Fe_4 molecules are partially oriented on the surface (the featureless XNLD of Fe_4C_9 can be found in Supplementary Fig. 8).

The preferential orientation was confirmed by periodic density functional theory (DFT) calculations performed on the chemisorbed configuration of a Fe_4C_5 molecule on an unreconstructed Au(111) surface (Fig. 3). In fact, the relaxed geometry indicates that the idealized three-fold molecular axis, which coincides with the easy magnetization axis, is restrained to form an angle smaller than 35° with the normal to the gold surface.

We simulated the observed XNLD signal within the framework of the ligand field multiplet approach (for details, see Methods and Supplementary Methods). Following previous calculations on the XAS and XMCD spectra of Fe_4, the cubic crystal field splitting (known as $10D_q$) for both the central and the peripheral ions was set to 1.5 eV (ref. 13), and the distortions from octahedral symmetry were adjusted to reproduce the magnetic anisotropies of the two sites, which are known from previous studies[21]. In particular, the easy-axis anisotropy of the central
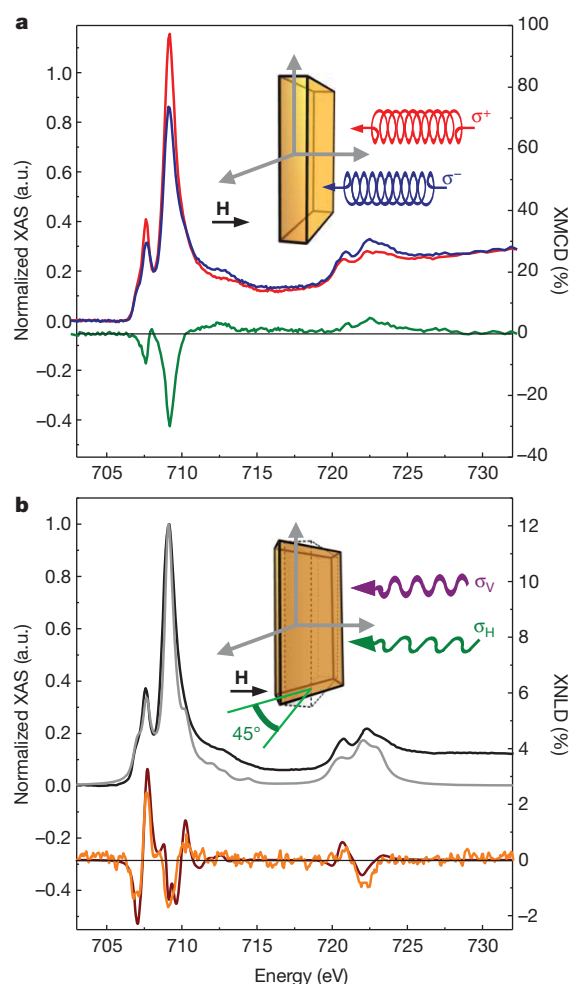
Figure 2 | X-ray absorption and dichroic spectra. a, Circularly polarized XAS (red, σ^+; blue, σ^-) and XMCD (green) spectra of the Fe_4C_5 monolayer at the iron $L_{2,3}$ edges ($T = 650(50)$ mK, $H = 30$ kOe). a.u., arbitrary units. **b**, Experimental (black) and calculated (grey) isotropic XAS ($1/3\sigma_V + 2/3\sigma_H$) together with the experimental (orange) and calculated (dark red) XNLD spectra ($\sigma_V - \sigma_H$), normalized to the isotropic signal ($T = 10(1)$ K, $H = 20$ kOe). As shown in the insets, the photon propagation vector is collinear with the applied field (**H**) and lies either along the surface normal (in XMCD) or at 45° to it (in XNLD).

iron ion along the molecular axis was treated by assigning to the low-symmetry crystal field parameters the values $D_\tau = 0.046$ eV and $D_\sigma = 0$ (see ref. 22 for precise definitions of D_q, D_τ and D_σ). The anisotropic contribution of the peripheral ions, in which the hard axis is parallel to the molecular plane, has been reproduced by setting $D_\tau = -0.069$ eV and $D_\sigma = 0$. Finally, we assumed an axial anisotropy for the monolayer along the surface normal (that is, the absence of any lateral ordering) and we averaged over all molecular orientations while constraining the molecular axis to lie within 35° from the surface normal. The model contains no adjustable parameters, so the good agreement between the experimental and calculated XNLD spectra (Fig. 2b) hints at a preferential orientation of the Fe_4 molecules with their easy magnetization axes perpendicular to the surface. The slight overestimation of the XNLD signal resulting from our calculations can be attributed to a small fraction of molecules having a more tilted adsorption geometry.

Given these indications, we recorded XMCD-detected hysteresis loops to study not only the static magnetic properties but also the spin dynamics of the monolayer. Figure 4a displays the hysteresis loops recorded at $T = 650(50)$ mK with a field sweeping rate of 20.0 Oe s^{-1}.

The opening of the hysteresis is apparent and confirms that a shorter spacer unit (in Fe_4C_5 relative to Fe_4C_9) does not destructively affect
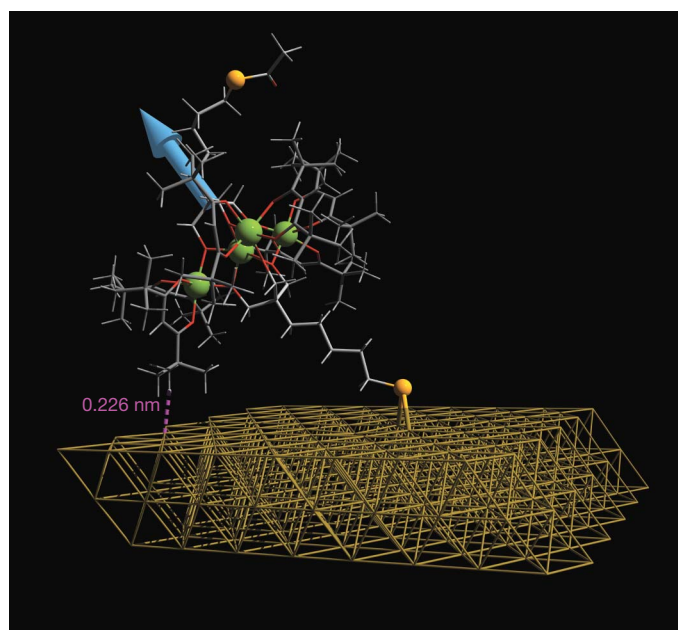
Figure 3 | Periodic DFT-optimized structure of a Fe$_4$C$_5$ cluster on an unreconstructed Au(111) surface. Green and yellow spheres indicate iron and sulphur atoms, respectively. The angle between the molecular axis, which corresponds to the easy magnetization axis (blue arrow), and the normal to the surface is limited to ~35° by the gold–hydrogen interaction (in purple) between an equatorial DPM ligand and the Au(111) surface.

0.226 nm

SMM behaviour at surfaces. However, the shape of the hysteresis shows evidence of resonant quantum tunnelling of the magnetization (QTM)[5,6], that is, the presence of accelerations in the dynamics when the longitudinal component of the field is $H_z = D/g\mu_B$ (where μ_B is the Bohr magneton and g is the gyromagnetic factor). At this field the $m = S$ and $m = -S + 1$ states are quasi-degenerate and admixed by transverse magnetic anisotropy (Fig. 4c). More importantly, by varying the angle, θ_H, between the magnetic field and the normal to the surface (Fig. 4d), a significant change in the hysteresis loop is observed. In particular, the step associated with resonant QTM shifts to higher fields with increasing θ_H, providing direct proof that the molecules are

preferentially oriented with their easy magnetization axes perpendicular to the surface.

To gain a deeper insight into the angular dependence of the magnetic hysteresis, we carried out a microscopic simulation of the data in Fig. 4a. Owing to the discreteness of the magnetic levels, the out-of-equilibrium magnetization can be estimated from the evolution of the population of the eleven components of the $S = 5$ ground state (considering the total splitting of the $S = 5$ state, the excited spin states are sufficiently higher in energy to assume that the giant spin approximation is acceptable). The time evolution of the population, $p(t)$, of each of the eleven states follows the Markov process described by the master equation[4]

$$\frac{d}{dt}p_p(t) = \sum_q [\gamma_q^p p_q(t) - \gamma_p^q p_p(t)] \qquad (1)$$

where γ_p^q is the probability of transition from the pth state of the multiplet to the qth (Methods). We note that for pure axial symmetry, the spin eigenfunctions are simply the eigenstates of $\hat{S}_z$, whereas in the presence of transverse magnetic anisotropy or transverse field they are linear combinations of these eigenstates (Methods). Therefore, our treatment allows us to take into account the resonant QTM because the spin eigenstates are substantially localized in one potential well far from resonance but are significantly delocalized close to level crossings[4]. Hence, the transition probabilities show discontinuities at the resonance fields, as depicted in Fig. 4c for the transition within the ground-state doublet in zero field. To reproduce numerically the magnetic hysteresis observed at subkelvin temperatures, we prepared the system in a strong magnetic field where only the ground state is populated. For each field step, we allowed the system to relax by recursively applying equation (1) after setting the unit of time and the number of cycles to match the experimental field sweeping rate. In our simulation, the parameter D, which describes the leading axial contribution to magnetic anisotropy, was held fixed at the value estimated from electron paramagnetic resonance spectroscopy. Such an assumption is realistic, considering that the tripodal ligands convey a considerable rigidity to molecular structure and that, within the field resolution of the XMCD experiment, the Fe$_4$C$_9$ monolayer showed the same D value as the bulk[7]. Many other parameters, such as second- and higher-order transverse anisotropy, the spin–phonon coupling, the speed of sound and so on, influence the hysteresis loop in a highly correlated fashion,
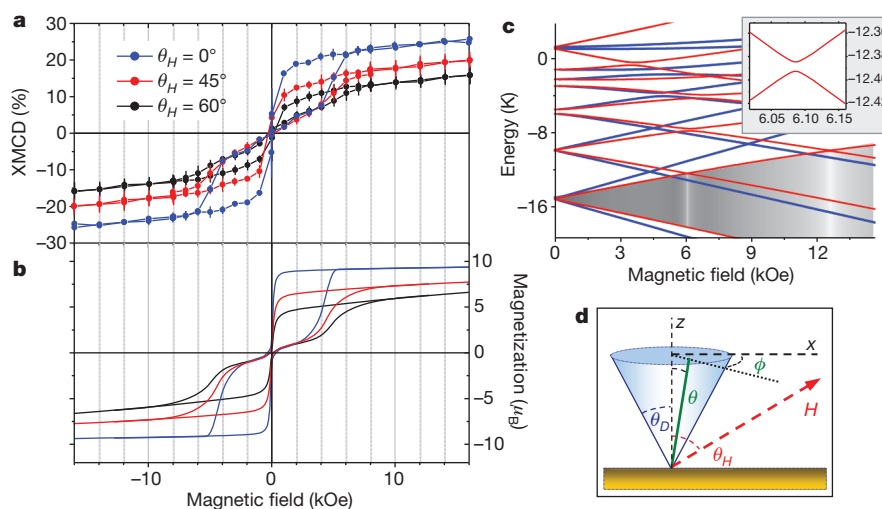
Figure 4 | Magnetic hysteresis loops. a, Angle-resolved hysteresis loops for the Fe$_4$C$_5$ monolayer obtained from the XMCD at 709.2 eV and $T = 650(50)$ mK. **b,** Calculated hysteresis loops. **c,** Zeeman diagrams calculated for $\theta_H = 0°$ (blue) and $45°$ (red) with $\theta_D = 0°$; the level crossings responsible for QTM shift up-field with increasing θ_H, the transition probability within the

zero-field ground doublet at $\theta_H = 45°$ (grey-scale filling) spans twelve orders of magnitude and a 10-mK tunnel splitting is found at the first level crossing (inset). **d,** Geometrical parameters describing the orientation of molecular axis (green line) on the surface.

with the result that they cannot be determined unambiguously. However, we satisfactorily reproduced the shape of the hysteresis by choosing a second-order transverse anisotropy with $|E/D| = 0.1$, which is slightly above the largest value exhibited by Fe_4 derivatives in the bulk phase[14] (0.07), and by adding the three-fold fourth-order term

$$\mathcal{H}_{\text{trans}} = \frac{1}{4} B_4^3 [\hat{S}_z(\hat{S}_+^3 + \hat{S}_-^3) + (\hat{S}_+^3 + \hat{S}_-^3)\hat{S}_z] \qquad (2)$$

with $B_4^3/k_B = 0.01$ K. The inclusion of the term in equation (2) is motivated by the idealized three-fold molecular symmetry and by the multispin origin of high-order anisotropy in SMMs[23]. It is worth stressing that a significant change in the efficiency of spin–phonon interactions is expected for the monolayer deposit as compared with the bulk phase. However, our low-temperature studies do not allow us to deconvolve its effect from the leading contribution of transverse anisotropy. Therefore, for simplicity we considered the spin–phonon coupling in the monolayer to be the same as in the bulk phase. Finally, dipolar interactions were neglected because the average distance between Fe_4 centres estimated from scanning tunnelling microscope images was ~5 nm and the resulting dipolar shifts in the resonance fields were expected to be well below our field resolution. In analogy to the procedure used in the analysis of XNLD spectra, we averaged the magnetic hysteresis over all possible orientations of the molecular axis, as described by the polar angles θ and ϕ, inside a cone defined by $\theta_D = 35°$ (Fig. 4d), in accordance with DFT calculations.

The results, shown in Fig. 4b, are in very good agreement with the experimental data at the three θ_H values explored and confirm that the proposed model accurately describes the grafting of Fe_4C_5 onto gold. The magnetization observed for $\theta_H = 0°$ shows a weak increase above 5 kOe whereas the calculated one is largely flat. We can attribute this increase to a small fraction of molecules that are not oriented inside the cone, because they are either physisorbed or grafted to gold adatoms.

In conclusion, our subkelvin investigation with synchrotron radiation, together with an exhaustive theoretical analysis, has shown that the characteristic field dependence of the relaxation rate due to resonant QTM is maintained when SMMs are wired to a gold surface. The steps in the hysteresis loops shift as a function of the applied field direction, as expected for a partially oriented monolayer and in accordance with the preferred adsorption geometry determined using *ab initio* calculations. The self-assembly process we used is thus suitable for obtaining arrays of intact and fully functional SMMs chemically grafted to gold. Orientation control is of relevance for any application of SMMs in single-molecule spintronic devices[8,24] and can be achieved by a rational chemical approach. From another point of view, it is interesting to note that SMMs and QTM effects, having constituted a milestone in the history of spin[25], are now proving useful in learning how complex matter organizes at the nanoscale.

METHODS SUMMARY

We synthesized crystalline Fe_4C_5 as described in Supplementary Methods. The monolayer was prepared as reported elsewhere[7]. Low-temperature XAS/XMCD investigations were carried out at the Swiss Light Source, using the TBT-XMCD endstation equipped with a ^{3}He–^{4}He dilution set-up[26]. XNLD spectra were recorded at the Dragon-ID08 beamline of the European Synchrotron Radiation Facility (France).

We performed DFT calculations with the CP2K program package (http://cp2k.berlios.de and ref. 27). Norm-conserving Goedecker–Teter–Hutter pseudopotentials were used for all atomic species. In addition, plane-wave basis sets with an energy cut-off of 350 Ry were used. The cut-off value was estimated in a previous work and the density functional used was TPSS[28].

In the simulation of the XNLD spectra, which was based on the electric dipole approximation[29], we took into account the orientation distribution and the adopted experimental geometry (Supplementary Methods and Supplementary Fig. 9).

In the quantum transfer matrix simulation of the hysteresis loops, the phonon-induced transition probability between the $|\varphi_p\rangle$ and $|\varphi_q\rangle$ spin eigenstates appearing in equation (1) was approximated using the leading terms of $|\langle \varphi_p | \mathcal{H}_{\text{s–ph}} | \varphi_q \rangle|^2$ ($\mathcal{H}_{\text{s–ph}}$ indicates the spin–phonon interaction) and is given by

$$
\begin{aligned}
\gamma_q^p = & \frac{3}{\pi h^4 \rho c_s^5} \frac{(E_p - E_q)^3}{([e^{(E_p - E_q)/k_B T} - 1])} \\
& \times \tilde{D}^2 (|\langle \varphi_p | \hat{S}_+^2 | \varphi_q \rangle|^2 + |\langle \varphi_p | \hat{S}_-^2 | \varphi_q \rangle|^2 \\
& + |\langle \varphi_p | \{\hat{S}_+, \hat{S}_z\} | \varphi_q \rangle|^2 + |\langle \varphi_p | \{\hat{S}_-, \hat{S}_z\} | \varphi_q \rangle|^2)
\end{aligned} \qquad (3)
$$

where E_p and E_q are the energies of the pth and qth states, ρ is the density of the material, c_s is the speed of sound in the material (considered isotropic), $\hbar$ is Planck's constant divided by 2π, $\{ , \}$ denotes anticommutation and only a spin–phonon coupling coefficient, $\tilde{D}$, is used at this level of approximation. The coefficient $\tilde{D}$ is not expected to vary with the magnetic field; however, the transition probabilities are strongly field dependent in both the density of phonons and in the matrix elements appearing in equation (3). The eigenstates of the system are in fact a linear combination of eigenstates of $\hat{S}_z$, $|\varphi_q\rangle = \sum_{m=-S}^{S} \lambda_m^q |m\rangle$ and the coefficients λ_m^q vary significantly around the level crossings. The magnetization was then numerically evaluated as $M(t, H) = -\sum_q p_q(t) dE_q/dH$ and averaged over all possible orientations inside the distribution cone.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

1. Loth, S. *et al.* Controlling the state of quantum spins with electric currents. *Nature Phys.* **6,** 340–344 (2010).
2. Parks, J. J. *et al.* Mechanical control of spin states in spin-1 molecules and the underscreened Kondo effect. *Science* **328,** 1370–1373 (2010).
3. Sessoli, R., Gatteschi, D., Caneschi, A. & Novak, M. A. Magnetic bistability in a metal-ion cluster. *Nature* **365,** 141–143 (1993).
4. Gatteschi, D., Sessoli, R. & Villain, J. *Molecular Nanomagnets* (Oxford Univ. Press, 2006).
5. Friedman, J. R., Sarachik, M. P., Tejada, J. & Ziolo, R. Macroscopic measurement of resonant magnetization tunneling in high-spin molecules. *Phys. Rev. Lett.* **76,** 3830–3833 (1996).
6. Thomas, L. *et al.* Macroscopic quantum tunnelling of magnetization in a single crystal of nanomagnets. *Nature* **383,** 145–147 (1996).
7. Mannini, M. *et al.* Magnetic memory of a single-molecule quantum magnet wired to a gold surface. *Nature Mater.* **8,** 194–197 (2009).
8. Bogani, L. & Wernsdorfer, W. Molecular spintronics using single-molecule magnets. *Nature Mater.* **7,** 179–186 (2008).
9. Rocha, A. R. *et al.* Towards molecular spintronics. *Nature Mater.* **4,** 335–339 (2005).
10. Heersche, H. B. *et al.* Electron transport through single Mn_{12} molecular magnets. *Phys. Rev. Lett.* **96,** 206801 (2006).
11. Jo, M. H. *et al.* Signatures of molecular magnetism in single-molecule transport spectroscopy. *Nano Lett.* **6,** 2014–2020 (2006).
12. Mannini, M. *et al.* XAS and XMCD investigation of Mn_{12} monolayers on gold. *Chem. Eur. J.* **14,** 7530–7535 (2008).
13. Mannini, M. *et al.* X-ray magnetic circular dichroism picks out single-molecule magnets suitable for nanodevices. *Adv. Mater.* **21,** 167–171 (2009).
14. Gregoli, L. *et al.* Magnetostructural correlations in tetrairon(III) single-molecule magnets. *Chem. Eur. J.* **15,** 6456–6467 (2009).
15. Margheriti, L. *et al.* Thermal deposition of intact tetrairon(III) single-molecule magnets in high-vacuum conditions. *Small* **5,** 1460–1466 (2009).
16. Zyazin, A. S. *et al.* Electric field controlled magnetic anisotropy in a single molecule. *Nano Lett.* **10,** 3307–3311 (2010).
17. Pineider, F. *et al.* Deposition of intact tetrairon(III) single molecule magnet monolayers on gold: an STM, XPS, and ToF-SIMS investigation. *J. Mater. Chem.* **20,** 187–194 (2010).
18. Nakajima, R., Stohr, J. & Idzerda, Y. U. Electron-yield saturation effects in L-edge X-ray magnetic circular dichroism spectra of Fe, Co, and Ni. *Phys. Rev. B* **59,** 6421–6429 (1999).
19. Gambardella, P. *et al.* Supramolecular control of the magnetic anisotropy in two-dimensional high-spin Fe arrays at a metal interface. *Nature Mater.* **8,** 189–193 (2009).
20. Van der Laan, G., Schofield, P. F., Cressey, G. & Henderson, C. M. B. Natural linear dichroism at the Fe 2p absorption edge of gillespite. *Chem. Phys. Lett.* **252,** 272–276 (1996).
21. Tancini, E. *et al.* Slow magnetic relaxation from hard-axis metal ions in tetranuclear single-molecule magnets. *Chem. Eur. J.* **16,** 10482–10493 (2010).
22. Juhin, A. *et al.* X-ray linear dichroism in cubic compounds: the case of Cr^{3+} in $MgAl_2O_4$. *Phys. Rev. B* **78,** 195103 (2008).
23. Barra, A. L. *et al.* The origin of transverse anisotropy in axially symmetric single molecule magnets. *J. Am. Chem. Soc.* **129,** 10754–10762 (2007).
24. Misiorny, M., Weymann, I. & Barnas, J. Spin diode behavior in transport through single-molecule magnets. *Europhys. Lett.* **89,** 18003 (2010).
25. Ziemelis, K. (1996) Mesoscopic tunnelling of magnetization. *Milestone 22: Nature Milestones in Spin* ⟨http://www.nature.com/milestones/milespin/full/milespin22.html⟩ (2008).

26. Letard, I. *et al.* Remnant magnetization of Fe$_8$ high-spin molecules: X-ray magnetic circular dichroism at 300 mK. *J. Appl. Phys.* **101,** 113920 (2007).

27. VandeVondele, J. *et al.* Quickstep: fast and accurate density functional calculations using a mixed Gaussian and plane waves approach. *Comput. Phys. Comm.* **167,** 103–128 (2005).

28. Bencini, A., Rajaraman, G., Totti, F. & Tusa, M. Modeling thiols on Au(111): structural, thermodynamic, and magnetic properties of simple thiols and thiol-radicals. *Superlattices Microstruct.* **46,** 4–9 (2009).

29. Brouder, C. Angular dependence of X-ray absorption spectra. *J. Phys. Condens. Matter* **2,** 701–738 (1990).

Author Contributions R.S., A.C. and M.M. designed the project; C.D. and A.C. synthesized the SMMs; L.S., A.C., F.P. and R.S. carried out the bulk magnetic characterization and analyses; M.M. and F.P. prepared the monolayers and performed the STM characterization and analysis; M.M., F.P., Ph.S., E.O., L.J., J.C.C. and R.S. performed the XMCD and XNLD experiments; F.T. performed the DFT calculations; Ph.S. and M.-A.A. performed the LFM calculations; and R.S. simulated the hysteresis curves.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to R.S. (roberta.sessoli@unifi.it).

METHODS

Crystalline [Fe$_4$(L)$_2$(DPM)$_6$] (Fe$_4$C$_5$) was synthesized as described in Supplementary Methods. The monolayer was prepared as reported elsewhere[7].

XAS, XMCD and XNLD characterizations were obtained following the procedure we established for SMMs[12] to avoid radiation damaging of the monolayer deposits. In all these experiments, absorbance was measured in the total-electron-yield mode[13,18] to achieve the required surface sensitivity.

Low-temperature XAS and XMCD investigations were carried out at the Swiss Light Source using the TBT-XMCD endstation equipped with a ^{3}He–^{4}He dilution set-up[26]. The same set-up was used to record hysteresis curves. The field dependence of the dichroic signal at the energy of its maximum amplitude (709.2 eV) was monitored with respect to the pre-edge (704.0-eV) background signal, using the two undulators of the X11MA-SIM beamline with opposite polarizations rapidly tuned and detuned, respectively. Each hysteresis was normalized with respect to the isotropic contribution at high field for each field orientation. Error bars on hysteresis data were evaluated by averaging over six field cycles. XNLD spectra were recorded exploiting the extreme stability and speed of the Dragon-ID08 beamline of European Synchrotron Radiation Facility (France) by using linearly polarized light produced by a downstream undulator and recording a set of 48 spectra.

DFT calculations were performed with the CP2K program package (http://cp2k.berlios.de and ref. 27). Norm-conserving Goedecker–Teter–Hutter (GTH) pseudopotentials were used for all atomic species. A GTH double-ζ, polarized, molecularly optimized basis set was used for iron and light elements. A GTH double-ζ basis set was used for gold atoms. In addition, plane-wave basis sets with an energy cut-off of 350 Ry were used. The cut-off value was estimated from a previous work and the employed density functional was TPSS[28]. The validity of the CP2K package, as an accurate k-point-only approach, was confirmed by a systematic supercell method for the calculation of transition-metal surface properties. The unit cell, containing 240 gold atoms, was shaped to obtain an Au(111) surface with three layers when the periodic boundary conditions are imposed over an orthorhombic simulation cell. The first layer was left to relax with the cluster ($S = 10$ state) until an energy plateau was reached. Afterwards, only the cluster was left to relax up to the imposed convergence criteria: 8×10^{-6} Ha for the SCF energy and 9×10^{-4} Ha Å^{-1} (Ha rad^{-1}) for the energy gradient. The simulation cell was defined with x and y axes on the surface plane. We chose a cell size along the z axis of 40 Å and an interlayer distance of 2.35 Å.

In the simulation of the XNLD spectra, we assumed that for the iron L$_{2,3}$ edges the electric dipole approximation is appropriate to describe the interaction of X-ray photons with matter. Following Brouder's model[29], it can then be assumed that the pleochroism of the sample is limited to the dichroic case where all cross-sections for any type of linear polarization can be obtained as a linear combination of two independent cross-sections measured for photons with linear polarizations respectively parallel ($\sigma_{//}$) and perpendicular ($\sigma_\perp$) to the anisotropy axis. In the evaluation of molecular XNLD, the cubic crystal field splitting ($10D_q = 1.5$ eV), the nephelauxetic reduction parameter ($\kappa = 0.6$), the spin–orbit coupling constant for $3d$ electrons (60 meV) and the $2p$ core-hole (8.2 eV) were set following previously published calculations[13]. The distortion parameters D_τ and D_σ were adjusted so as to reproduce the magnetic anisotropies of the Fe(III) sites, which are known from previous studies[21]. The detailed procedure used to compute the XNLD signal from a monolayer sample, taking into account the orientation distribution and the adopted experimental geometry, can be found in Supplementary Methods and Supplementary Fig. 9.

In the quantum transfer matrix simulation of the hysteresis loops, the phonon-induced transition probability between the $|\varphi_p\rangle$ and $|\varphi_q\rangle$ spin eigenstates appearing in equation (1) was approximated using the leading terms of $|\langle\varphi_p|\mathcal{H}_{s-ph}|\varphi_q\rangle|^2$ ($\mathcal{H}_{s-ph}$ indicates the spin–phonon interaction) and is given by

$$
\begin{aligned}
\gamma_q^p = &\frac{3}{\pi\hbar^4\rho c_s^5}\frac{(E_p - E_q)^3}{([e^{(E_p-E_q)/k_BT} - 1])} \\
&\times \tilde{D}^2(|\langle\varphi_p|\hat{S}_+^2|\varphi_q\rangle|^2 + |\langle\varphi_p|\hat{S}_-^2|\varphi_q\rangle|^2 \\
&+ |\langle\varphi_p|\{\hat{S}_+,\hat{S}_z\}|\varphi_q\rangle|^2 + |\langle\varphi_p|\{\hat{S}_-,\hat{S}_z\}|\varphi_q\rangle|^2)
\end{aligned}
\tag{3}
$$

where E_p and E_q are the energies of the pth and qth states, ρ is the density of the material, c_s is the speed of sound in the material (considered isotropic), $\hbar$ is Planck's constant divided by 2π, { , } denotes anticommutation and only a spin–phonon coupling coefficient, $\tilde{D}$, is used at this level of approximation. The coefficient $\tilde{D}$ is not expected to vary with the magnetic field; however, the transition probabilities are strongly field dependent in both the density of phonons and in the matrix elements appearing in equation (3). The eigenstates of the system are in fact a linear combination of eigenstates of $\hat{S}_z$, $|\varphi_q\rangle = \sum_{m=-S}^{S}\lambda_m^q|m\rangle$ and the coefficients λ_m^q vary significantly around the level crossings. The magnetization was then numerically evaluated as $M(t, H) = -\sum_q p_q(t)dE_q/dH$ and averaged over all possible orientations inside the distribution cone.

LETTER

Structure and mechanism of the S component of a bacterial ECF transporter

Peng Zhang[1], Jiawei Wang[2] & Yigong Shi[3]

The energy-coupling factor (ECF) transporters, responsible for vitamin uptake in prokaryotes, are a unique family of membrane transporters[1,2]. Each ECF transporter contains a membrane-embedded, substrate-binding protein (known as the S component), an energy-coupling module that comprises two ATP-binding proteins (known as the A and A′ components) and a transmembrane protein (known as the T component). The structure and transport mechanism of the ECF family remain unknown. Here we report the crystal structure of RibU, the S component of the ECF-type riboflavin transporter from *Staphylococcus aureus* at 3.6-Å resolution. RibU contains six transmembrane segments, adopts a previously unreported transporter fold and contains a riboflavin molecule bound to the L1 loop and the periplasmic portion of transmembrane segments 4–6. Structural analysis reveals the essential ligand-binding residues, identifies the putative transport path and, with sequence alignment, uncovers conserved structural features and suggests potential mechanisms of action among the ECF transporters.

The ATP-binding cassette (ABC) transporters harness the energy of ATP hydrolysis to move substrate molecules across membrane. An importer of the ABC superfamily comprises two cytosolic ABC domains, two membrane-spanning domains and a periplasmic binding protein that specifically recognizes substrate (Fig. 1a). Structural investigations on the ABC transporters have revealed major insights into their function and mechanism of action[3–11]. Despite a similar organization (Fig. 1a), the ABC and ECF transporters have different organizational and functional properties. In contrast to the ABC importer, the S component of the ECF-type transporter is responsible for substrate binding (Fig. 1a) and there are cases where the S component alone is able to mediate high-capacity transport of substrate[12]. The S component does not exhibit sequence homology with any protein of known structure. RibU[13,14] in *Lactococcus lactis* and YpaA[15] in *Bacillus subtilis* are the S components of the ECF-type transporters for riboflavin, the essential precursor for flavin mononucleotide and flavin adenine dinucleotide.
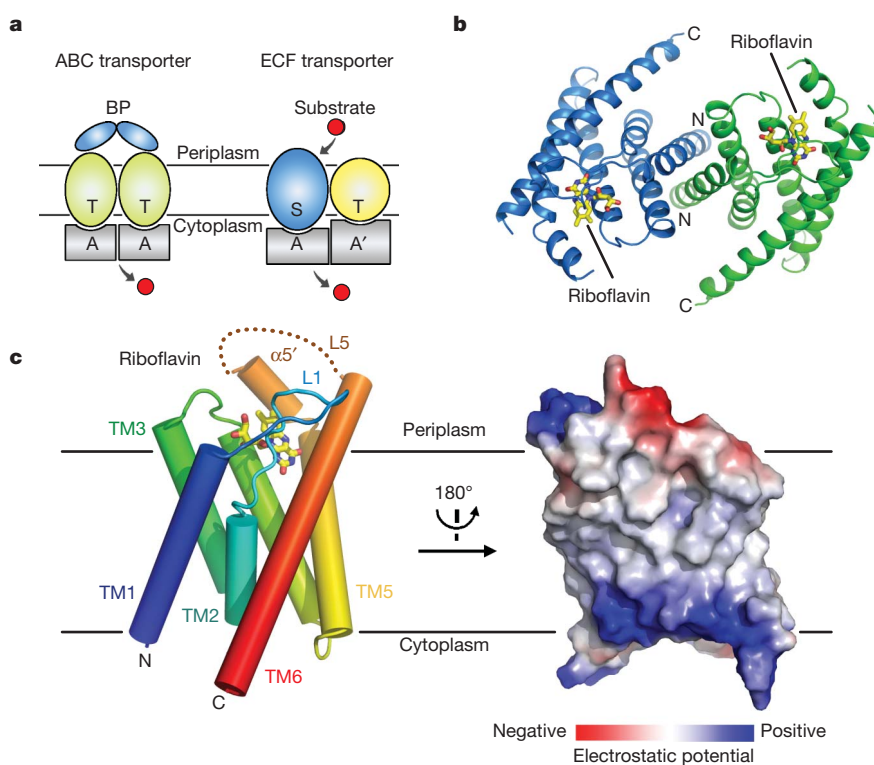
Figure 1 | The overall structure of RibU. a, The ECF transporter is distinct from other known ABC transporters. We show a schematic comparison between a representative ECF transporter and an ABC importer. Substrate is directly recognized by the S component of the ECF transporter, whereas substrate is bound to the binding protein in the ABC importer. **b,** Ribbon representation of a RibU homodimer in one asymmetric unit. Each RibU molecule modelled here and elsewhere in the figures contains amino acids 10–141 and 153–188. **c,** Structure of a RibU molecule in ribbon diagram (left) and surface electrostatic potential (right). RibU is positioned roughly perpendicular to the lipid membrane. All structural figures were prepared with PyMol[24].

[1]Department of Molecular Biology, Lewis Thomas Laboratory, Princeton University, Princeton, New Jersey 08544, USA. [2]State Key Laboratory of Biomembrane and Membrane Biotechnology, Center for Structural Biology, School of Life Sciences and School of Medicine, Tsinghua University, Beijing 100084, China. [3]Ministry of Education Protein Science Laboratory, Center for Structural Biology, School of Life Sciences and School of Medicine, Tsinghua University, Beijing 100084, China.

We cloned, expressed, purified and crystallized RibU from *S. aureus*. The presence of riboflavin in the recombinant RibU protein was suggested by the absorption spectrum[14] (Supplementary Fig. 1a) and confirmed by mass spectrometry. The identity and function of RibU were confirmed by both *in vitro* analysis, where RibU formed a stable complex with the corresponding T, A and A′ components from *S. aureus* (Supplementary Fig. 1b), and *in vivo* analysis, where all four components were required to support growth of riboflavin-auxotrophic *Escherichia coli* strains (Supplementary Fig. 1c). After numerous trials, we determined the structure by multi-wavelength anomalous dispersion at 3.6-Å resolution (Supplementary Table 1 and Supplementary Fig. 2).

There are two molecules of RibU in an asymmetric unit, arranged as a pseudo-symmetric dimer (Fig. 1b). At present, we have no evidence to support the biological relevance of the dimeric arrangement, which would result in a relatively short membrane-spanning distance of about 20 Å and burial of highly charged surface patches in the hydrophobic interior of the lipid membrane (Supplementary Fig. 3). For simplicity, we limit our discussion to one RibU molecule.

The overall structure of RibU resembles a cylinder with rugged ends (Fig. 1c). Assignment of RibU orientation in the membrane was facilitated by the observation that the carboxy (C) terminus of YpaA resides in the cytoplasm[15]. The outer surface of the cylinder is predominantly hydrophobic, consistent with its membrane-buried nature. By contrast, the cytoplasmic and periplasmic faces are enriched with charged amino acids (Supplementary Fig. 4). RibU comprises six transmembrane segments, not five as previously reported[13,15], where transmembrane segments 2 and 3 were predicted to be a single transmembrane segment. Each RibU protein contains a riboflavin molecule, which is bound on the periplasmic side about 5 Å into the predicted membrane surface (Fig. 1c).

Except transmembrane segment 2, which only contains a short, 11-residue helix, each of the other five transmembrane segments contains a continuous α-helix that probably traverses the entire lipid membrane (Fig. 1c). The intervening sequences between transmembrane segments 2 and 3, transmembrane segments 3 and 4, and transmembrane segments 4 and 5 are relatively short (Fig. 2a). An extended loop between transmembrane segments 1 and 2 (the L1 loop) contains 17 amino acids, nine of which are highly conserved among representative RibU homologues (Supplementary Fig. 5). The L1 loop hovers above the substrate-binding site, suggesting an important role (Fig. 1c).

Despite a similar reliance on the ATP-binding domains for substrate transport, the fold of RibU is markedly different from those of the ABC transporters (Supplementary Fig. 6). A search of the Protein Data Bank using DALI[16] failed to identify any entry that is structurally homologous

to RibU over its entire six transmembrane segments. In particular, no structure of any membrane transporter was found to be similar to RibU. Among the proteins that exhibit limited structural similarity with RibU, five of the top seven entries are derived from particulate methane monooxygenase[17,18], a membrane-bound metalloenzyme. Transmembrane segments 1–5 of RibU can be superimposed with chain F of particulate methane monooxygenase[18] with a root mean squared deviation of 3.3 Å over 124 aligned Cα atoms (Fig. 2b and Supplementary Fig. 7).

The amino-acid sequences of RibU homologues from eight bacterial species share a high degree of pairwise sequence identity (Supplementary Fig. 5), suggesting structural conservation. We reasoned that the highly conserved amino acids among these RibU homologues may be functionally important. To examine this possibility, we mapped the conserved amino acids onto the structure of RibU (Fig. 3 and Supplementary Fig. 8). The outer surface of RibU only contains a small proportion of the conserved amino acids (Fig. 3a), whereas most of the highly conserved residues are clustered in the interior (Fig. 3b). Notably, four invariant amino acids are located around the substrate-binding pocket. In addition, the conserved amino acids also map to the interior of the cylinder-shaped RibU molecule, populating from the substrate-binding pocket to the cytoplasmic side. These amino acids appear to define the putative transport path for substrate.

The riboflavin-binding pocket measures approximately 15 Å in width and 8 Å in thickness (Fig. 4a); it is capped by the L1 loop on the periplasmic side. Despite the moderate resolution, the experimental electron density for riboflavin was clearly visible (Supplementary Fig. 9a). Nonetheless, we chose to model riboflavin after all protein atoms were in place. Riboflavin contains a ribityl side chain, with four hydroxyl groups, and an aromatic ring, which is hydrophobic on one end and polar on the other. Analysis of the structural features of RibU (Supplementary Fig. 9b) suggests only one way of orienting riboflavin into the binding pocket.

Riboflavin is recognized by relatively conserved amino acids from loop L1 and transmembrane segments 4–6, through both hydrogen bonds and van der Waals interactions (Fig. 4b, c). The non-polar portion of the riboflavin ring is nestled in a hydrophobic cage, involving 13 amino acids. These include Tyr 41/Leu 42 on the L1 loop, Val 83/Gly 84/Ala 87/Asn 88/Ala 91 on transmembrane segment 4, Leu 127/Val 134/Leu 135/Leu 138 on transmembrane segment 5 and the small helix α5′, and Ile 160/Phe 163 on transmembrane segment 6 (Fig. 4b and Supplementary Fig. 10). In addition, there may be eight H bonds between riboflavin and the conserved residues from RibU (Fig. 4c and Supplementary Fig. 10). In particular, Asn 131 and Asn 164, both of which are invariant among the RibU homologues, may mediate direct
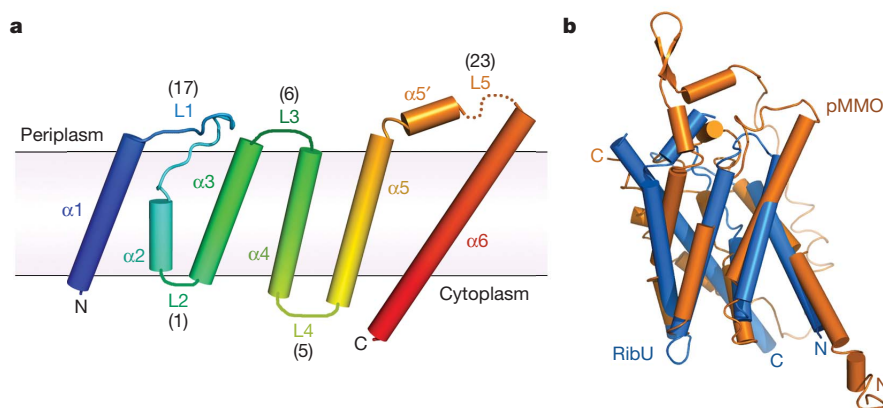
Figure 2 | Sequence alignment and structural fold of RibU. a, Membrane topology of RibU. The lengths of the loops connecting neighbouring transmembrane segments are indicated in parentheses. The periplasmic loop between transmembrane segments 5 and 6 (the L5 loop) has 23 amino acids, of which seven form a short α-helix (α5′). Eleven amino acids in this loop

(residues 142–152) have no electron density and are presumably disordered in the crystals. **b**, Structural overlay of RibU (blue) with chain F of the particulate methane monooxygenase (orange) from *Methylosinus trichosporium OB3B* (orange, Protein Data Bank accession code 3CHX[18]).
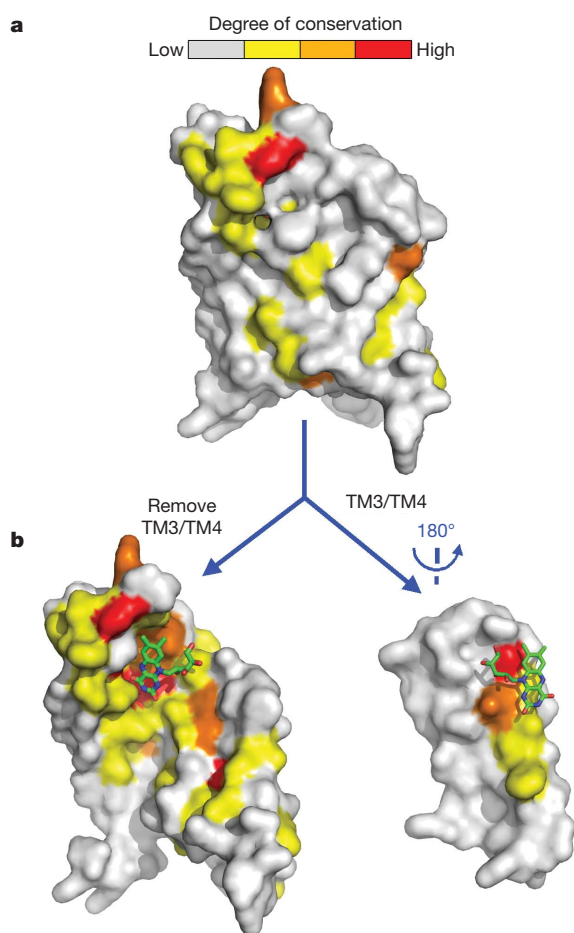
Figure 3 | Conserved amino acids map to the binding site and transport path of riboflavin. a, Mapping of conserved amino acids onto the structure of RibU. Based on the sequence alignment of 12 RibU homologues, residues that are conserved in seven to nine and 10 to 11 bacterial species are coloured yellow and orange, respectively. Invariant residues are highlighted in red. A surface representation is shown here. **b,** The riboflavin-binding site is enriched by highly conserved amino acids. The RibU molecule is split into two portions, transmembrane segment 3/transmembrane segment 4 and the rest, to reveal the location of the highly conserved amino acids.

H bonds to riboflavin. Similarly, Thr 43 and Lys 167, both from conserved positions in RibU homologues, are also within H-bond distances of riboflavin.

The interactions between riboflavin and RibU are extensive, consistent with the reported binding affinity of approximately 0.6 nM between riboflavin and the *L. lactis* RibU[14]. By contrast, flavin mononucleotide interacted with RibU with a moderate affinity of 36 nM, whereas flavin adenine dinucleotide exhibited no measurable binding[14]. These observations are nicely explained by our structure-based modelling analysis. Because the ribityl side chain of riboflavin is positioned towards the small periplasmic opening of the substrate-binding pocket, the phosphate group of flavin mononucleotide, but not the adenine dinucleotide of flavin adenine dinucleotide, can be tolerated by minor conformational shifts of the surrounding residues (Supplementary Fig. 11).

The relatively simple topology of RibU reveals tantalizing clues about how substrate might be imported from periplasm to the cytoplasm. RibU can be thought of having two structural modules: transmembrane segments 1–3 and 4–6 (Fig. 1c). Riboflavin is bound between these two modules, with the L1 loop coming from the transmembrane segment 1–3 module. Under this arrangement, riboflavin is probably transported through the central line of RibU surrounded by
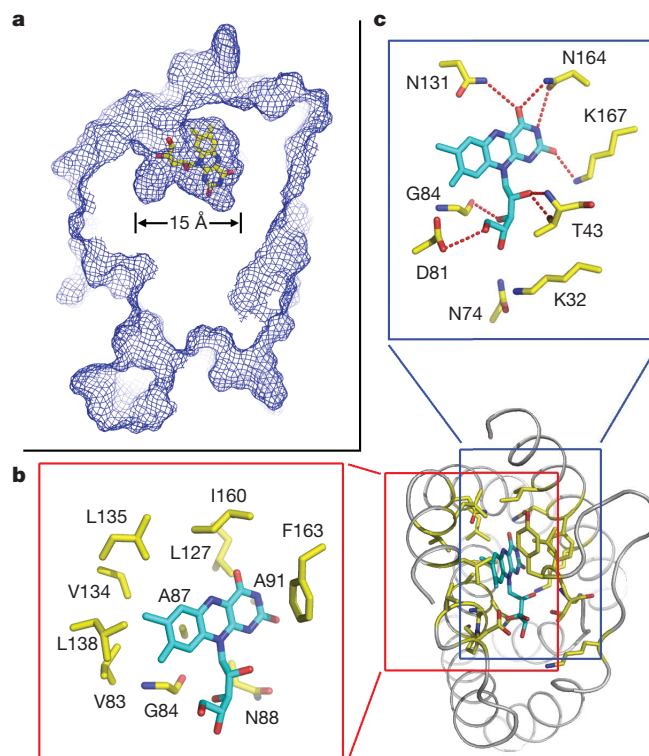
Figure 4 | Recognition of riboflavin by RibU. a, Riboflavin is bound at the periplasmic side of RibU. We show a slice of RibU to highlight the riboflavin-binding pocket. The surface of RibU is represented by blue mesh. **b,** Riboflavin is nestled in a hydrophobic pocket formed by 13 conserved amino acids. Except Tyr 41 and Leu 42, which are removed to present a clear view here, all other amino acids are shown. **c,** Riboflavin is recognized by multiple hydrogen bonds. Potential hydrogen bonds are represented by red dashed lines.

transmembrane segments 1–3 and 4–6. This analysis gives rise to a speculative working model (Supplementary Fig. 12). In this model, the L1 loop is thought to serve as a gate. Upon binding to riboflavin, L1 may close down to interact with the substrate molecule. Then, facilitated by the T–A–A′ components as a result of ATP hydrolysis, the transmembrane segment 1–3 module may be moved away from transmembrane segments 4–6, allowing the protein to adopt a transient, inward-open conformation. Such changes may lead to disruption of interactions with riboflavin, allowing it to be released into the cytoplasm. The ADP-bound state probably resets the transport system.

Sequence alignment of RibU with representative transporters, such as those for folate, thiamine precursor and cobalamin precursor, revealed a pattern of conservation that is closely associated with the transmembrane segments of RibU (Supplementary Fig. 13). This observation suggests that the S components of at least some ECF transporters may contain six transmembrane segments, have a similar structure and adopt the same membrane topology. Notably, this conclusion may not apply to other S components such as the bipartite proteins CbiMN and NikMN[2]. Sequence alignment of the transporters for folate and cobalamin precursor identified candidate sequences that may be responsible for binding to their respective ligands (Supplementary Fig. 14).

The ECF transporters are classified into two groups[1]. Group I transporters are encoded by linked genes, which encode the S component and an S-specific A–T module. Group II transporters have a common A–T module that is shared by up to 12 different S components in the same bacterial species[1]. The RibU transporter from *S. aureus* belongs to group II. Surprisingly, sequence alignment of the group II S components from the same bacterial species failed to uncover any conserved sequence feature, suggesting that interaction with the shared

A–T module may only entail short and/or degenerate motifs. This notion is consistent with the identification of two short but functionally essential Arg motifs in the T components of the ECF transporters[19].

For the cytoplasmic A components to use the energy of ATP hydrolysis for substrate transport, the S component should contain a cytoplasmic motif that mediates this interaction. In addition, the candidate motif must be associated with one or more of the following cytoplasmic elements: the amino (N) and C termini, the L2 loop and the L4 loop. This analysis, and the divergent nature of the N terminus and the observation that the L2 loop contains only a single amino acid, Gly, suggests that the L4 loop and/or the C-terminal sequences are probably responsible for binding to the A components. Supporting this conjecture, the group II S components appear to have a stretch of positively charged amino acids in the L4 loop and the C terminus (Supplementary Fig. 15).

Structural elucidation of RibU represents the first of many required steps towards mechanistic understanding of the ECF transporters. At present, we have little information about how the S component interacts with the energy-coupling module or how ATP hydrolysis by the A components facilitates the transport of substrate. Answers to these questions require systematic biochemical and structural investigation.

METHODS SUMMARY

RibU was overexpressed in *E. coli*, purified to homogeneity and crystallized by the hanging-drop vapour-diffusion method. All data were collected at the X29 beamline of the Brookhaven National Laboratory and processed with HKL2000 (ref. 20). The crystals belong to the space group $P2_12_12_1$ with unit cell dimensions of $a = 50.4$ Å, $b = 94.2$ Å, $c = 115.4$ Å. Additional processing was performed using programs from the CCP4 suite[21]. Multiwavelength anomalous diffraction phasing was done using Phenix AutoSol. The initial model was built using the incorporated Resolve in Phenix. Additional missing residues in the auto-built model were manually added in COOT[22]. The final model was refined using PHENIX[23].

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

1. Rodionov, D. A. *et al.* A novel class of modular transporters for vitamins in prokaryotes. *J. Bacteriol.* **191,** 42–51 (2009).
2. Rodionov, D. A., Hebbeln, P., Gelfand, M. S. & Eitinger, T. Comparative and functional genomic analysis of prokaryotic nickel and cobalt uptake transporters: evidence for a novel group of ATP-binding cassette transporters. *J. Bacteriol.* **188,** 317–327 (2006).
3. Hollenstein, K., Dawson, R. J. & Locher, K. P. Structure and mechanism of ABC transporter proteins. *Curr. Opin. Struct. Biol.* **17,** 412–418 (2007).
4. Rees, D. C., Johnson, E. & Lewinson, O. ABC transporters: the power to change. *Nature Rev. Mol. Cell Biol.* **10,** 218–227 (2009).
5. Davidson, A. L., Dassa, E., Orelle, C. & Chen, J. Structure, function, and evolution of bacterial ATP-binding cassette systems. *Microbiol. Mol. Biol. Rev.* **72,** 317–364 (2008).
6. Oldham, M. L., Davidson, A. L. & Chen, J. Structural insights into ABC transporter mechanism. *Curr. Opin. Struct. Biol.* **18,** 726–733 (2008).
7. Linton, K. J. Structure and function of ABC transporters. *Physiology* **22,** 122–130 (2007).
8. Dawson, R. J., Hollenstein, K. & Locher, K. P. Uptake or extrusion: crystal structures of full ABC transporters suggest a common mechanism. *Mol. Microbiol.* **65,** 250–257 (2007).
9. Davidson, A. L. & Maloney, P. C. ABC transporters: how small machines do a big job. *Trends Microbiol.* **15,** 448–455 (2007).
10. Locher, K. P. Review. Structure and mechanism of ATP-binding cassette transporters. *Phil. Trans. R. Soc. B* **364,** 239–245 (2009).
11. Jones, P. M., O'Mara, M. L. & George, A. M. ABC transporters: a riddle wrapped in a mystery inside an enigma. *Trends Biochem. Sci.* **34,** 520–531 (2009).
12. Hebbeln, P., Rodionov, D. A., Alfandega, A. & Eitinger, T. Biotin uptake in prokaryotes by solute transporters with an optional ATP-binding cassette-containing module. *Proc. Natl Acad. Sci. USA* **104,** 2909–2914 (2007).
13. Burgess, C. M. *et al.* The riboflavin transporter RibU in *Lactococcus lactis*: molecular characterization of gene expression and the transport mechanism. *J. Bacteriol.* **188,** 2752–2760 (2006).
14. Duurkens, R. H., Tol, M. B., Geertsma, E. R., Permentier, H. P. & Slotboom, D. J. Flavin binding to the high affinity riboflavin transporter RibU. *J. Biol. Chem.* **282,** 10380–10386 (2007).
15. Vogl, C. *et al.* Characterization of riboflavin (vitamin B2) transport proteins from *Bacillus subtilis* and *Corynebacterium glutamicum*. *J. Bacteriol.* **189,** 7367–7375 (2007).
16. Holm, L. & Sander, C. Protein structure comparison by alignment of distance matrices. *J. Mol. Biol.* **233,** 123–138 (1993).
17. Lieberman, R. L. & Rosenzweig, A. C. Crystal structure of a membrane-bound metalloenzyme that catalyses the biological oxidation of methane. *Nature* **434,** 177–182 (2005).
18. Hakemian, A. S. *et al.* The metal centers of particulate methane monooxygenase from *Methylosinus trichosporium* OB3b. *Biochemistry* **47,** 6793–6801 (2008).
19. Neubauer, O. *et al.* Two essential arginine residues in the T components of energy-coupling factor transporters. *J. Bacteriol.* **191,** 6482–6488 (2009).
20. Otwinowski, Z. & Minor, W. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276,** 307–326 (1997).
21. Collaborative Computational Project, N. The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. D* **50,** 760–763 (1994).
22. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D* **60,** 2126–2132 (2004).
23. Adams, P. D. *et al.* PHENIX: building new software for automated crystallographic structure determination. *Acta Crystallogr. D* **58,** 1948–1954 (2002).
24. DeLano, W. L. The PyMOL molecular graphics system. <http://www.pymol.org> (2002).

METHODS

Protein preparation. We tested four different bacterial species (*S. aureus*, *Thermofilum pendens*, *Thermoanaerobacter italicus* and *B. subtilis*) by cloning, expressing and purifying the S component of the ECF transporter for riboflavin. Crystals of *B. subtilis* RibU were beautiful but failed to diffract X-rays. We focused on the RibU protein from *S. aureus*. The RibU coding sequence from *S. aureus* was chemically synthesized (Genescript), subcloned into pET15b (Novagen) and over-expressed in *E. coli* BL21(DE3) by 0.5 mM β-D-thiogalactopyranoside (IPTG) at A_{600} of about 0.8. After 14 h at 37 °C, the cells were collected, homogenized in the buffer containing 20 mM Tris-HCl, pH 8.0 and 100 mM NaCl, and lysed using a French Press with two passes at 15–20,000 p.s.i. Cell debris was removed by centrifugation. The supernatant was collected and applied to ultracentrifugation at 150,000*g* for 1 h. Membrane fraction was incubated with 2% (w/v) nonyl-β-D-glucopyranoside (β-NG, Anatrace) for 2 h at 4 °C. After another ultracentrifugation step at 150,000*g* for 30 min, the supernatant was loaded to an Ni^{2+}-NTA affinity column (Qiagen). The protein was eluted from the affinity resin by 20 mM Tris, pH 8.0, 500 mM imidazole and 0.4% β-NG and concentrated to around 10 mg ml^{-1} before further purification by gel filtration (Superdex-200, GE Healthcare). The buffer for gel filtration contained 20 mM Tris (pH 8.0), 100 mM NaCl and 0.4% β-NG. The peak fraction was collected and concentrated to approximately 8 mg ml^{-1} for crystallization.

Formation of the RibU-T-A-A′ complex. Genes encoding the four putative components of riboflavin ECF transporter in *S. aureus*, RibU, T, A and A′ were subcloned into pQlink and pACYCDuet vectors to obtain two expression plasmids: pQlink-A′-A and pACYCDuet-RibU-T. The gene identity and predicted molecular masses are as follows: RibU, 161509653, 21.1 kDa; A, 15925211, 32.9 kDa; A′, 15925212, 30.0 kDa; T, 15925210, 30.8 kDa. A tag of six histidine residues was added at the C terminus of the A component and the N terminus of RibU. These two plasmids were transformed separately or co-transformed into *E. coli* BL21(DE3). The A and A′ components could be co-expressed in a stable complex and were co-purified by Ni2+-NTA affinity resin (Qiagen), followed by gel filtration chromatography (Superdex 200, GE Healthcare). By contrast, co-expression of RibU with the T component only led to expression and purification of RibU alone. The T component could only be expressed and purified in the presence of all three other components (A, A′ and RibU).

Co-expression of all four components was achieved by co-transforming *E. coli* BL21(DE3) with the plasmids pQlink-A′-A and pACYCDuet-RibU-T (with 6×His tag at the C terminus of A and N terminus of RibU). The quaternary complex RibU–T–A–A′ was purified in three sequential steps. First, the complex was co-purified by Ni2+-NTA affinity resin (Qiagen) and eluted with 500 mM imidazole. Second, the eluted proteins were fractionated by anion exchange chromatography (Source-15Q, GE Healthcare) using a linear gradient of 0–500 mM NaCl in 20 mM Tris buffer (pH 8.0). The quaternary complex RibU–T–A–A′ stayed together on the anion exchange column and was co-eluted in the same fractions. Third, the RibU–T–A–A′ complex was concentrated and further purified by gel filtration chromatography (Superdex-200, GE Healthcare). The gel filtration buffer contained 20 mM Tris, pH 8.0, 0.1 M NaCl, 0.04% DDM.

***In vivo* experiments.** Two *E. coli* riboflavin-auxotrophic strains, ribB11 mutant BSV11 (F⁻ glnV44(AS) λ⁻ mcrA rfbC1 endA1 ribB11:Tn5 spoT1 thi-1 mcrB hsdR29) and ribA13 mutant BSV13 (F⁻ glnV44(AS) λ⁻ mcrA rfbC1 endA1 ribA13:Tn5 spoT1 thi-1 mcrB hsdR29), were obtained from the Yale Coli Genetic Stock Center (numbers 6991 and 6992). These two mutant strains were unable to synthesize riboflavin owing to disruption of the riboflavin biosynthesis pathway[25]. The two mutants are unable to grow in regular lysogeny broth medium (which contains an unknown amount of riboflavin) but can grow after addition of

20 mg l^{-1} riboflavin. The lysogeny broth plates were prepared in the presence of 1 mM IPTG, four concentrations of additional riboflavin (0, 0.5, 2.5 and 12.5 mg l^{-1}) and appropriate antibiotics (100 mg ml^{-1} ampicillin, 50 mg ml^{-1} kanamycin and 34 mg ml^{-1} chloramphenicol). To make the strains inducible by IPTG, these two mutants were lysogenized with a λ DE3 Lysogenization Kit (Novagen). The mutant *E. coli* (DE3) strains were transformed individually with plasmids expressing RibU, T, RibU+T, A + A′ and RibU + T + A + A′, and cultured in lysogeny broth with additional 20 mg l^{-1} riboflavin. The control culture was not transformed by any plasmid. The overnight culture was diluted to an A_{600} of 0.1. An equal volume of the diluted culture (10 μl) was dispensed onto the lysogeny broth plates, occupying the upper half of each plate. The lower half of each plate was used to streak the culture from the upper half. The plates were incubated at 37 °C overnight.

Crystallization. Crystals were grown at 20 °C by the hanging-drop vapour-diffusion method. Several RibU homologues from other bacterial species were cloned, purified and attempted in crystallizations. Only RibU protein from *S. aureus* gave rise to crystals of reasonable diffraction (RibU hereafter). Large yellow-coloured crystals of RibU were obtained in many conditions containing polyethylene glycol. However, none of the crystals diffracted X-rays beyond 5 Å at the synchrotron. Polyethylene glycols with low molecular masses were found to support crystals of better diffraction. The best crystals, which diffracted X-rays to about 5 Å, were generated in 29% polyethylene glycol 550 MME, 0.1 M Tris-HCl, pH 8.4. To improve the diffraction quality further, we screened secondary detergent from a detergent screening kit (Hampton Research) and available detergents from Anatrace, each with varying ratios of protein to detergent. Addition of octyl-maltoside fluorinated to the protein with a ratio of 1:1 to 2:1 led to improvement of the diffraction limit from 5 to 4 Å. The best crystal diffracted X-rays to 3.6 Å at the X29 beamline of Brookhaven National Laboratory. The Se-Met protein crystals used for MAD phasing were obtained in a similar manner and diffracted X-rays to 3.8-Å resolution. Both native and Se-Met crystals were directly flash frozen in a cold nitrogen stream at 100 K.

Data collection and structure determination. All data sets were collected at the X29 beamline of the Brookhaven National Laboratory and processed with HKL2000 (ref. 20). The crystals belong to the space group $P2_12_12_1$ with unit cell dimensions of $a = 50.4$ Å, $b = 94.2$ Å, $c = 115.4$ Å. Additional processing was performed using programs from the CCP4 suite[21]. Data collection statistics are summarized in Supplementary Table 1. MAD phasing was done using Phenix AutoSol; six selenium sites were found, four of which were above 5 standard deviations, corresponding to Met 20 and Met 123 in the two molecules, and two of which were above 3 standard deviations, corresponding to Met 9 and Met 79 in one of the two molecules. The initial model was built using the incorporated Resolve in Phenix. Additional missing residues in the auto-built model were manually added in COOT[22] with the aid of the map sharpening utility. The final model in the $P2_12_12_1$ space group was refined using PHENIX[23] with tight restraints, including non-crystallographic symmetry, experimental phases and α-helix main-chain hydrogen-bond restraints.

Mass spectrometry identification. The purified RibU protein is yellow, suggesting the presence of a bound substrate molecule. To identify the yellow-coloured 'substrate', 10 μl purified RibU was denatured at 95 °C for 2 min, followed by addition of 40 μl water and centrifugation at 16,000*g* for 5 min. The yellow-coloured supernatant (40 μl) was then diluted with additional 360 μl water. This sample was subjected to liquid chromatography–mass spectrometry. Riboflavin (5 μM) was used as a positive control.

25. Bandrin, S. V., Rabinovich, P. M. & Stepanov, A. I. Three linkage groups of the genes of riboflavin biosynthesis in *Escherichia coli*. *Genetika* **19,** 1419–1925 (1983).

LETTER

Bottom–up effects of plant diversity on multitrophic interactions in a biodiversity experiment

Christoph Scherber[1,2], Nico Eisenhauer[3,4], Wolfgang W. Weisser[2], Bernhard Schmid[5], Winfried Voigt[2], Markus Fischer[6,7], Ernst-Detlef Schulze[8], Christiane Roscher[2,9], Alexandra Weigelt[2,10], Eric Allan[2], Holger Beßler[11], Michael Bonkowski[12], Nina Buchmann[13], François Buscot[14], Lars W. Clement[2], Anne Ebeling[1,2], Christof Engels[11], Stefan Halle[2], Ilona Kertscher[2], Alexandra-Maria Klein[2,15], Robert Koller[12], Stephan König[14], Esther Kowalski[2], Volker Kummer[7], Annely Kuu[16], Markus Lange[2], Dirk Lauterbach[2], Cornelius Middelhoff[8], Varvara D. Migunova[17], Alexandru Milcu[18], Ramona Müller[2], Stephan Partsch[3], Jana S. Petermann[5,19], Carsten Renker[14,20], Tanja Rottstock[7], Alexander Sabais[3], Stefan Scheu[3], Jens Schumacher[8,21], Vicky M. Temperton[8,22] & Teja Tscharntke[1]

Biodiversity is rapidly declining[1], and this may negatively affect ecosystem processes[2], including economically important ecosystem services[3]. Previous studies have shown that biodiversity has positive effects on organisms and processes[4] across trophic levels[5]. However, only a few studies have so far incorporated an explicit food-web perspective[6]. In an eight-year biodiversity experiment, we studied an unprecedented range of above- and below-ground organisms and multitrophic interactions. A multitrophic data set originating from a single long-term experiment allows mechanistic insights that would not be gained from meta-analysis of different experiments. Here we show that plant diversity effects dampen with increasing trophic level and degree of omnivory. This was true both for abundance and species richness of organisms. Furthermore, we present comprehensive above-ground/below-ground biodiversity food webs. Both above ground and below ground, herbivores responded more strongly to changes in plant diversity than did carnivores or omnivores. Density and richness of carnivorous taxa was independent of vegetation structure. Below-ground responses to plant diversity were consistently weaker than above-ground responses. Responses to increasing plant diversity were generally positive, but were negative for biological invasion, pathogen infestation and hyperparasitism. Our results suggest that plant diversity has strong bottom-up effects on multitrophic interaction networks, with particularly strong effects on lower trophic levels. Effects on higher trophic levels are indirectly mediated through bottom-up trophic cascades.

The loss of biodiversity from terrestrial ecosystems has been shown to affect ecosystem properties, such as primary productivity[7], nutrient cycling[8] and trophic interactions[2]. In recent biodiversity experiments, focal organism groups (usually plants[7]) were used to establish gradients in species richness, and biodiversity effects were then measured at one or a few trophic levels[5,9]. Traditionally, studies have focused on the effects of horizontal biodiversity loss, that is, loss of species within a single trophic level[10]. Biodiversity loss at a given trophic level has been predicted to affect the abundance, biomass and resource use of that trophic level[5]. However, horizontal species loss may also affect other

trophic levels, organism groups and processes, and, hence, vertical species loss and the associated multitrophic structure of ecosystems[10]. For example, declines in plant species richness may cause losses to herbivores, true predators, parasitoids, hyperparasitoids and omnivores, and may also alter mutualistic interactions such as pollination[11] or mycorrhizal association[4]. Overall, there is an increasing awareness that the network nature of ecological systems needs to be incorporated into studies of biodiversity–ecosystem functioning[12].

Recent meta-analyses[4,5] and experiments at individual study sites[13,14] have shown plant diversity effects on a wide range of different groups of organisms, including primary producers, first- and second-order consumers, detritivores, fungal diseases and mycorrhizae. Additional studies have addressed components of the below-ground subsystem and their linkages with above-ground biota[15]. However, interpretation and progress has been clouded by differences in study systems and by a general lack of an overarching theory incorporating both trophic and non-trophic interactions as well as direct and indirect interactions[16,17]. So far, subcomponents of food webs have often been studied in isolation, for example primary producers, the decomposer subsystem[18], soil nematodes[19], soil microbes, plant pathogenic fungi[20], above-ground invertebrates[13], pollinators[21] and so on. Here we present data from one of the most comprehensive biodiversity experiments so far, and show that diversity effects on higher trophic levels are mostly indirect and mediated through bottom-up trophic cascades. We use structural equation modelling approaches to develop comprehensive above-ground/below-ground biodiversity food webs. Finally, we link our results to recent interaction web models and provide explicit parameter estimates that can be used in future modelling exercises.

We experimentally manipulated plant species and functional group richness in 82 sown grassland plots (Methods), and recorded abundances and species richness of all relevant organism groups and biotic interactions between 2002 and 2009 (Supplementary Table 1). All data were analysed on a standardized scale[22] from zero to one and the relationship between plant species richness and the different response variables was modelled using a power function[18] to allow comparisons and extrapolation to other systems (see Supplementary Table 1 and

[1]Georg-August-University Göttingen, Department of Crop Sciences, Agroecology, Grisebachstrasse 6, 37077 Göttingen, Germany. [2]Friedrich Schiller University of Jena, Institute of Ecology, Dornburger Strasse 159, 07743 Jena, Germany. [3]Georg-August-University Göttingen, J. F. Blumenbach Institute of Zoology and Anthropology, Berliner Strasse 28, 37073 Göttingen, Germany. [4]University of Minnesota, Department of Forest Resources, 115 Green Hall, 1530 Cleveland Avenue North, St Paul, Minnesota 55108, USA. [5]Institute of Environmental Sciences, University of Zürich, Winterthurerstrasse 190, 8057 Zürich, Switzerland. [6]University of Bern, Plant Sciences, Altenbergrain 21, 3013 Bern, Switzerland. [7]University of Potsdam, Institute of Biochemistry and Biology, Maulbeerallee 1, 14469 Potsdam, Germany. [8]Max-Planck-Institute for Biogeochemistry, Hans-Knöll-Strasse 10, 07745 Jena, Germany. [9]UFZ Helmholtz Centre for Environmental Research, Department of Community Ecology, Theodor-Lieser-Strasse 4, 06120 Halle, Germany. [10]University of Leipzig, Institut für Biologie I, Johannisallee 21-23, 04103 Leipzig, Germany. [11]Humboldt University, Department of Plant Nutrition, Albrecht-Thaer-Weg 4, 14195 Berlin, Germany. [12]University of Cologne, Cologne Biocenter, Department of Terrestrial Ecology, Zülpicher Strasse 47b, 50674 Cologne, Germany. [13]ETH Zurich, Institute of Plant, Animal and Agroecosystem Sciences, LFW C56, Universitätstrasse 2, 8092 Zurich, Switzerland. [14]UFZ Helmholtz Centre for Environmental Research, Department of Soil Ecology, Theodor-Lieser-Straße 4, 06102 Halle, Germany. [15]Leuphana University of Lueneburg, Institute of Ecology and Environmental Chemistry, Scharnhorststraße 1, 21335 Lüneburg, Germany. [16]Tallinn University of Technology, Tartu College, Puiestee 78, 51008 Tartu, Estonia. [17]K. I. Skryabin All-Russian Institute of Helminthology, Bolschaya Tcheryomuschkinskaya Street 28, 117218 Moscow, Russia. [18]NERC Centre for Population Biology, Division of Biology, Imperial College London, Silwood Park Campus, Ascot SL5 7PY, United Kingdom. [19]University of British Columbia, Department of Zoology, 6270 University Boulevard, Vancouver, British Columbia V6T 1Z4, Canada. [20]Mainz Museum of Natural History, Reichklarastraße 10, 55116 Mainz, Germany. [21]Friedrich Schiller University of Jena, Institute of Stochastics, Ernst-Abbe-Platz 2, 07743 Jena, Germany. [22]Forschungszentrum Jülich GmbH, Phytosphere Institute (ICG-3), 52425 Jülich, Germany.

Supplementary Fig. 3 for untransformed data). Analyses consisted of three steps. First, every response variable was analysed separately using a common set of linear, saturating and exponential models with untransformed plant species richness as the main explanatory variable. The presence of legumes and grasses and the number of plant functional groups were fitted as additional covariates. Variance heterogeneity was modelled using variance functions. Model selection was based on the Akaike information criterion for small sample sizes (AICc). Then, for parsimony, models were refitted using a power function. This allowed comparisons between the abundance and species richness of herbivores, carnivores and all other functional groups. Finally, multivariate techniques (multivariate linear models and structural equation models) were used to account for non-independence of variables measured on the same field plots.

Plant species richness had highly significant overall effects on the abundances of other organisms ($T_{PB} = 0.56$ (Pillai–Bartlett trace), approximately F-distributed with $F_{10,37} = 4.741$, $P < 0.001$; Fig. 1a, c), the species richness of other organism groups ($T_{PB} = 0.788$, approx. $F_{9,38} = 15.69$, $P < 0.001$; Fig. 1b, d) and on trophic interactions ($T_{PB} = 0.733$, approx. $F_{10,22} = 6.04$, $P < 0.001$; Supplementary Fig. 1; see Supplementary Methods for definitions of interactions). The abundance and species richness of organisms and biotic interactions were affected in broadly similar ways by changes in plant species richness (Fig. 1 and Supplementary Fig. 1).

Model selection using the complete range of linear, saturating and exponential models (Supplementary Tables 2 and 3) showed that 90% of all relationships could be approximated by a power model of the form $y = a + bS^z$ (ref. 18), where the exponent z can take any real value (in particular zero and one as special cases). Only five out of 38 organism groups declined with plant species richness (abundances of hyperparasitoids, fungivorous nematodes and mites, and abundance and

species richness of plant invaders; Supplementary Table 4). Responses of the below-ground subsystem were consistently smaller (average power model exponent of 0.11) than above-ground responses (exponent of 0.14).

Although most responses were saturating, closer inspection (Supplementary Table 5a–c) revealed consistent differences between the responses of herbivores, carnivores, omnivores and other trophic groups that are likely to reflect a general pattern (Fig. 2): with increasing trophic distance and for omnivores, species richness effects dampened—as indicated by the magnitude of the exponent of the common power function (Supplementary Table 4). This effect was found both for organism abundances and organism species richness, both above and below ground, and it was further supported by structural equation models (Fig. 3 and Supplementary Tables 6–10). Together, these findings indicate that species richness effects are generally dampened along trophic cascades.

If plant species richness acts on other organisms along trophic cascades, and plant species richness is the only experimentally manipulated variable, then the simplest conceptual model in our case is a bottom-up model of plant species richness effects; that is, plant species richness effects are passed from one trophic level to the next. Several authors have suggested such a 'bottom-up template' perspective for terrestrial food webs[23]. Both decomposers and predators have long
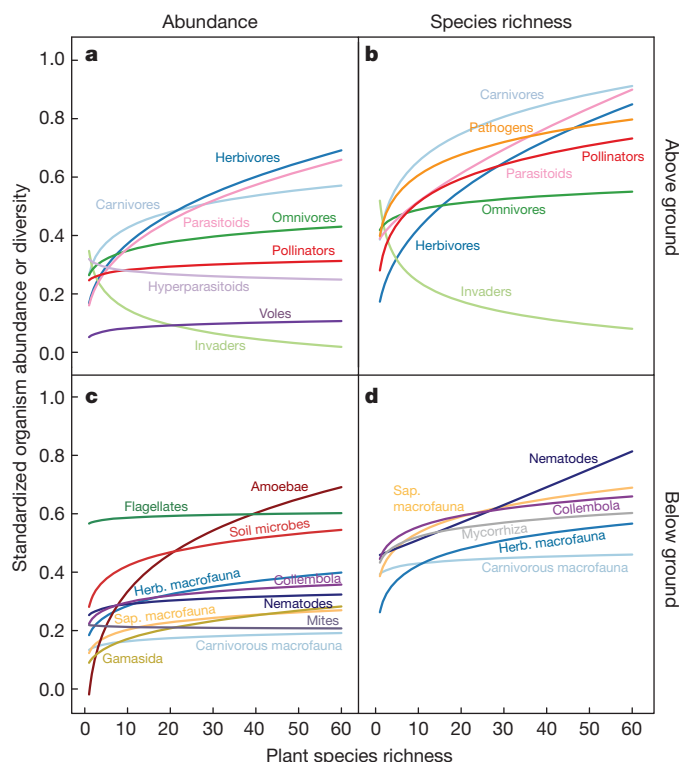
Figure 1 | Effects of plant species richness on above- and below-ground organisms in temperate grassland. a, b, Abundance (a) and species richness (b) of above-ground organisms. c, d, Abundance (c) and species richness (d) of below-ground organisms. All response variables scaled to [0, 1]. Every curve is fitted using a power function with covariates (Methods). Identical colours in each pair of panels indicate identical groups of organisms. For sample sizes, see Supplementary Table 1. Herb., herbivorous; Sap., saprophagous.
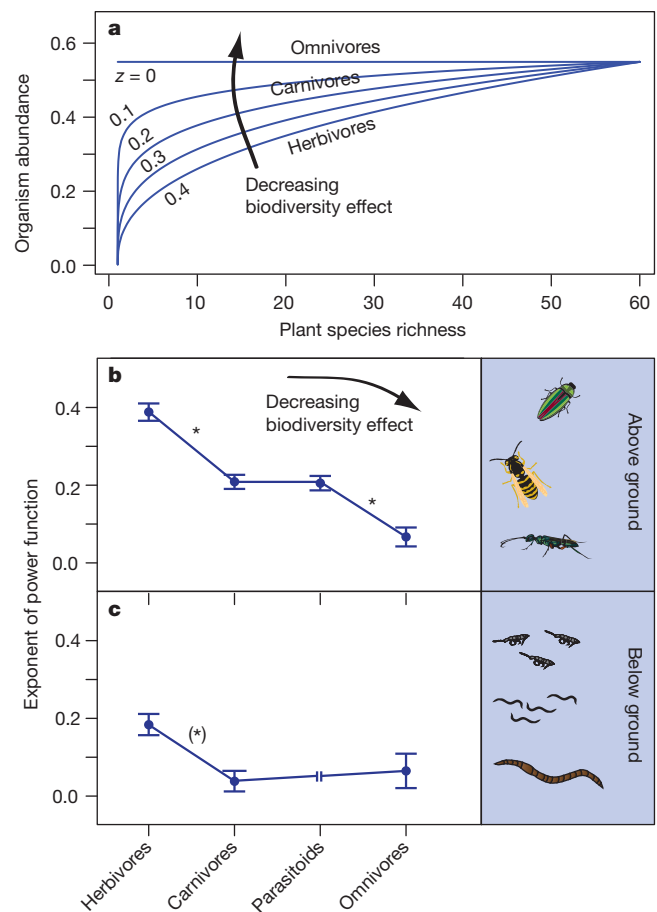
Figure 2 | Dampening of plant species richness effects with increasing trophic level. a, Conceptual figure showing how different values of z may influence biodiversity effects (x axis shows example range of 1–60 plant species). b, Estimates of z for above-ground herbivores, carnivores, parasitoids and omnivores. c, As in b, but for below-ground organisms. The y axes in b and c show estimated exponents of power functions fitted to data scaled to [0, 1]. Significant differences in z values are indicated by asterisks (*$P < 0.05$, $N = 50$ for above-ground organisms; (*)$P = 0.06$, $N = 82$ for below-ground organisms). Estimates are model predictions ± s.e.
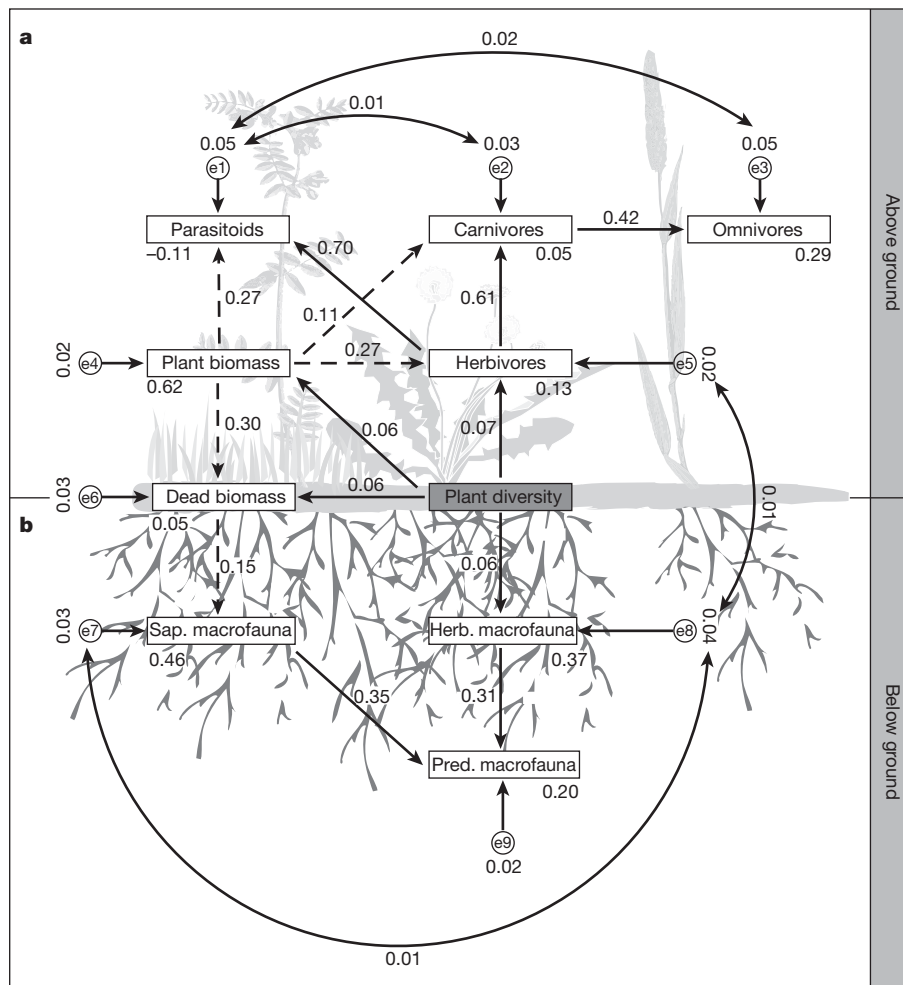
---

with carnivore biomass indirectly controlled by plant and herbivore biomass, and top-down control of herbivores by carnivores. Structural equation models are a powerful tool for detecting such mutual dependencies, greatly enhancing our understanding of biodiversity effects in multitrophic systems. Overall, our results from a wide variety of organism groups provide strong support for a prominent role of plant species richness (rather than productivity or other covariates) in shaping multitrophic interactions.

Our results present the intriguing possibility that the effects of the species richness of one trophic level on others decrease with trophic distance. This hypothesis merits exploration by means of experimental manipulations of species numbers on other trophic levels. Because even an experiment as large as ours (82 plots) limits how many variables can reasonably be included in a multiple regression or structural equation model, future studies should be designed explicitly with a particular network of trophic interactions in mind. These studies could also be combinations of observational and experimental approaches.

We scaled all response variables to allow us to seek generalizations across different types of organism and trophic levels, but note that unscaled analyses might offer other types of insight. We also note that detailed collection of data at the level of each individual species, although prohibitively time consuming in a broad survey such as ours, is also likely to offer added insight. Our study should therefore be seen as a starting point rather than as an end point for further analyses of other data sets.

We have shown that the consequences of biodiversity loss are consistently negative for most organism groups and interactions, with particularly far-reaching feedback effects on basal trophic levels. Below-ground organisms will be less affected by biodiversity change (or will respond more slowly) than above-ground ones. Changes in plant species richness will affect neighbouring trophic levels and cascade up to higher trophic levels. Exponents of power functions ($y = bS^z$) will decline with trophic level. Our results highlight the importance of a diverse resource base[28] for trophic interactions in terrestrial ecosystems.

METHODS SUMMARY

Experimental design. In a 10-ha former arable field near Jena (Germany), we controlled the number of plant species, functional groups and plant functional identity in 82 plots, each 20 m × 20 m, in a randomized block design[29]. Plots were seeded in May 2002 with 1, 2, 4, 8, 16 or 60 perennial grassland plant species, with 16, 16, 16, 16, 14 and 4 replicates, respectively. Plot compositions were randomly chosen from 60 plant species typical for local *Arrhenatherum* grasslands. Plots were maintained by mowing, weeding and herbicide applications.

Ecosystem variables. Sown and realized plant species richness were highly correlated (2006: Spearman's rank correlation coefficient, 0.995; $t = 91.94$; 80 degrees of freedom; $P < 2.2 \times 10^{-16}$); hence, sown richness was used for analysis. Above-ground invertebrates were collected on $N = 50$ plots using pitfall traps and suction sampling. Below-ground macro- and mesofauna were extracted from Kempson soil cores. Special sampling protocols were used for microorganisms (fungi, bacteria). Decomposition was measured using litter bags. Flower visitation was a count of pollinator visits. Parasitism was measured using a trap-nest technique. Hyper-parasitism was measured from aphid mummy counts in 6.25-m² replicate plots. Pathogen damage above ground and herbivory were estimated visually. Plant invasion was a count of the numbers of an invader plant species per unit area. Microbial biomass was measured using glucose as an artificial substrate. A full description is available in the Supplementary Methods.

Statistics. Explanatory variables in linear models were block, plant species richness, plant functional group richness, and grass and legume presence. Nonlinear models contained plant species richness, with legume and grass presence and functional richness as covariates. Models were simplified and compared using AICc. To test for differences between slopes, multivariate linear models were constructed, and orthogonal contrasts were used to test linear hypotheses. Structural equation models were fitted to test specific hypotheses on causal relationships.

1. Sala, O. E. et al. Global biodiversity scenarios for year 2100. *Science* **287**, 1770–1774 (2000).
2. Hooper, D. U. et al. Effects of biodiversity on ecosystem functioning: a consensus of current knowledge. *Ecol. Monogr.* **75**, 3–35 (2005).
3. Myers, N. Environmental services of biodiversity. *Proc. Natl Acad. Sci. USA* **93**, 2764–2769 (1996).
4. Balvanera, P. et al. Quantifying the evidence for biodiversity effects on ecosystem functioning and services. *Ecol. Lett.* **9**, 1146–1156 (2006).
5. Cardinale, B. J. et al. Effects of biodiversity on the functioning of trophic groups and ecosystems. *Nature* **443**, 989–992 (2006).
6. Cardinale, B. et al. in *Biodiversity, Ecosystem Functioning, and Human Wellbeing: an Ecological and Economic Perspective* (eds S. Naeem et al.) 105–120 (Oxford Univ. Press, 2009).
7. Hector, A. et al. Plant diversity and productivity experiments in European grasslands. *Science* **286**, 1123–1127 (1999).
8. Hooper, D. U. & Vitousek, P. M. Effects of plant composition and diversity on nutrient cycling. *Ecol. Monogr.* **68**, 121–149 (1998).
9. Petchey, O. L. et al. Species loss and the structure and functioning of multitrophic aquatic systems. *Oikos* **104**, 467–478 (2004).
10. Duffy, J. E. et al. The functional role of biodiversity in ecosystems: incorporating trophic complexity. *Ecol. Lett.* **10**, 522–538 (2007).
11. Biesmeijer, J. C. et al. Parallel declines in pollinators and insect-pollinated plants in Britain and the Netherlands. *Science* **313**, 351–354 (2006).
12. Tylianakis, J. M., Tscharntke, T. & Lewis, O. T. Habitat modification alters the structure of tropical host-parasitoid food webs. *Nature* **445**, 202–205 (2007).
13. Haddad, N. M. et al. Plant species loss decreases arthropod diversity and shifts trophic structure. *Ecol. Lett.* **12**, 1029–1039 (2009).
14. Koricheva, J., Mulder, C. P. H., Schmid, B., Joshi, J. & Huss-Dannell, K. Numerical responses of different trophic groups of invertebrates to manipulations of plant diversity in grasslands. *Oecologia* **125**, 271–282 (2000).
15. Wardle, D. A. & Van der Putten, W. H. in *Biodiversity and Ecosystem Functioning* (eds Loreau, M., Naeem, S. & Inchausti, P.) 155–168 (Oxford Univ. Press, 2002).
16. Goudard, A. & Loreau, M. Nontrophic interactions, biodiversity, and ecosystem functioning: an interaction web model. *Am. Nat.* **171**, 91–106 (2008).
17. Ogushi, T., Craig, T. P. & Price, P. W. (eds) *Ecological Communities: Plant Mediation in Indirect Interaction Webs* (Cambridge Univ. Press, 2007).
18. Srivastava, D. S. et al. Diversity has stronger top-down than bottom-up effects on decomposition. *Ecology* **90**, 1073–1083 (2009).
19. De Deyn, G. B. Raaijmakers, C. E., van Ruijen, J., Berendse, F. & van der Putten, W. H. Plant species identity and diversity effects on different trophic levels of nematodes in the soil food web. *Oikos* **106**, 576–586 (2004).
20. Mitchell, C. E., Tilman, D. & Groth, J. V. Effects of grassland plant species diversity, abundance, and composition on foliar fungal disease. *Ecology* **83**, 1713–1726 (2002).
21. Ghazoul, J. Floral diversity and the facilitation of pollination. *J. Ecol.* **94**, 295–304 (2006).
22. Legendre, P. & Legendre, L. *Numerical Ecology* (Elsevier, 1998).
23. Hunter, M. D. & Price, P. W. Playing chutes and ladders: heterogeneity and the relative roles of bottom-up and top-down forces in natural communities. *Ecology* **73**, 724–732 (1992).
24. Hairston, N. G., Smith, F. E. & Slobodkin, L. B. Community structure, population control, and competition. *Am. Nat.* **94**, 421–425 (1960).
25. Schmid, B. et al. in *Biodiversity, Ecosystem Functioning & Human Wellbeing* (eds Naeem, S. et al.) Ch. 2, 14–29 (Oxford Univ. Press, 2010).
26. Eisenhauer, N. et al. Plant diversity effects on soil microorganisms support the singular hypothesis. *Ecology* **91**, 485–496 (2010).
27. Grace, J. B. *Structural Equation Modeling and Natural Systems* (Cambridge Univ. Press, 2006).
28. Price, P. W. The resource-based organization of communities. *Biotropica* **24**, 273–282 (1992).
29. Roscher, C. et al. The role of biodiversity for element cycling and trophic interactions: an experimental approach in a grassland community. *Basic Appl. Ecol.* **5**, 107–121 (2004).

Author Contributions C.S., C.M. and J.S. assembled all data. N.E. contributed data for the below-ground food webs. C.S. performed all analyses and wrote the manuscript. C.S. and T.T. are the principal authors, N.E. to A.W. are listed in descending order of the importance of their contributions, and the remaining authors contributed equally. All authors contributed their own data sets, discussed the analysis and results, and commented on the manuscript text.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to C.S. (christoph.scherber@agr.uni-goettingen.de).

LETTER

Climate-driven population divergence in sex-determining systems

Ido Pen[1], Tobias Uller[2], Barbara Feldmeyer[1]†, Anna Harts[1], Geoffrey M. While[3] & Erik Wapstra[3]

Sex determination is a fundamental biological process, yet its mechanisms are remarkably diverse[1,2]. In vertebrates, sex can be determined by inherited genetic factors or by the temperature experienced during embryonic development[2,3]. However, the evolutionary causes of this diversity remain unknown. Here we show that live-bearing lizards at different climatic extremes of the species' distribution differ in their sex-determining mechanisms, with temperature-dependent sex determination in lowlands and genotypic sex determination in highlands. A theoretical model parameterized with field data accurately predicts this divergence in sex-determining systems and the consequence thereof for variation in cohort sex ratios among years. Furthermore, we show that divergent natural selection on sex determination across altitudes is caused by climatic effects on lizard life history and variation in the magnitude of between-year temperature fluctuations. Our results establish an adaptive explanation for intra-specific divergence in sex-determining systems driven by phenotypic plasticity and ecological selection, thereby providing a unifying framework for integrating the developmental, ecological and evolutionary basis for variation in vertebrate sex determination.

Vertebrates exhibit both genotypic (GSD) and temperature-dependent sex determination (TSD)[1,2]. The latter is particularly common in reptiles and both systems can co-occur within taxonomic families[3]. In addition, some species show elements of both genotypic and environmental sex determination within populations[4,5]. The causes of repeated evolutionary shifts between GSD and TSD and the origin and maintenance of mixed systems are two of the greatest unsolved problems in sex determination research[1–8]. The main reasons that diversity in reptilian sex determination has remained an enigma has been a failure empirically to link incubation temperature to ecological conditions promoting TSD and to establish theoretically that those conditions are sufficient to drive evolutionary shifts in sex-determining systems[3,7]. Here we provide both kinds of support using evolutionary models parameterized with field data to show how climatic effects on lizard life history generate evolutionary divergence in sex-determining systems via natural selection on sex ratios.

Environment-dependent sex determination can be favoured over genotypic sex determination when there are sex-specific fitness effects of environmental conditions experienced during or after the sex-determining period[9]. Temperature has a strong effect on the rate of embryonic development in ectotherm animals, with relatively cool conditions resulting in delayed birth or hatching. Sex differences in the fitness consequences of timing of birth could therefore favour integration of temperature-dependent developmental processes and gonad differentiation to ensure a match between offspring sex and birth date[10,11]. As a result, spatial or temporal variation in the strength of sex-specific selection on birth date, and therefore on TSD, may explain rapid evolutionary divergence in sex determination between populations or species[10–12].

The snow skink, *Niveoscincus ocellatus,* is a small live-bearing lizard occurring along a 1,200-m altitudinal, and climatic, gradient from sea level to highland regions throughout Tasmania[13]. Sex determination is affected by maternal basking opportunity in lowland skinks, analogous to temperature-dependent sex determination in egg-laying reptiles[14]. Thermal conditions representative of a cool year delays birth and result in an overproduction of male offspring whereas thermal conditions representative of warm years result in early birth and a small female bias (Fig. 1a). However, experimental manipulation of female thermal opportunity during gestation (a common garden experiment) reveals that sex determination in highland populations is not affected by temperature (Fig. 1b). This difference in sex-determining systems has consequences for sex ratios at the population level, with a negative correlation between the cohort sex ratio and annual temperature in lowland, but not highland, populations ($r = -0.84$, $P = 0.017$, $N = 7$ and $r = -0.20$, $P = 0.65$, $N = 7$, respectively; slopes differ significantly between populations, $F_{1,10} = 12.8$, $P = 0.005$).

Earlier birth for females may be adaptive because birth date affects opportunity for growth until maturity, which is more important in female than in male snow skinks as a result of differences in selection on body size[14–16]. However, climatic conditions vary substantially across altitudes and the cooler conditions in highland regions induce several changes in lizard life history. High-altitude populations have a shorter activity season, more synchronized birth, slower growth and delayed age at maturity compared to lowland populations[13,17]. Birth date is therefore a relatively unimportant predictor of the onset of maturity and reproductive output at high altitudes (Fig. 2a). Specifically, at low altitudes early-born females have about 50% higher lifetime fitness than late-born females, whereas at high altitudes the effect of birth date on
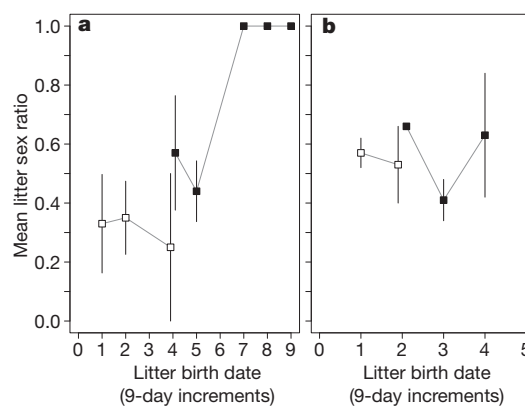
Figure 1 | Experimental effects of thermal conditions on sex ratio and birth date. Sex ratio = male/(male + female). Poor thermal condition during gestation (filled squares) results in delayed birth compared to good thermal condition (open squares), with a corresponding significant effect on offspring sex in lowland (**a**) but not highland (**b**) females. Error bars are s.e.m. Logistic regression with the proportion of males as a dependent variable and treatment and birth date (measured in days from birth) as predictors: birth date for lowland population $\chi^2 = 20.66$, $P = 0.0001$, $N_{females} = 13$, 18 and for highland population, $\chi^2 = 0.15$, $P = 0.70$, $N_{females} = 31$, 24.

[1]Theoretical Biology Group, University of Groningen, PO Box 14, 9750 AA Haren, the Netherlands. [2]Edward Grey Institute, Department of Zoology, University of Oxford, South Parks Road, Oxford OX1 3PS, UK. [3]School of Zoology, Private Bag 5, University of Tasmania, Hobart 7001, Tasmania, Australia. †Present address: Biodiversity and Climate Research Centre (BiK-F), Siesmayerstrasse 70A, D-60325 Frankfurt and Main, Germany.
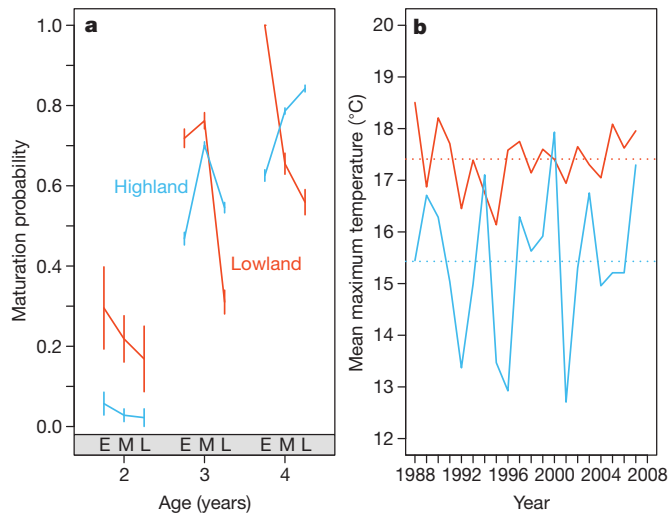
Figure 2 | Life-history and temperature differences between lowland and highland populations of *N. ocellatus*. a, Probability of maturation (±s.e.m.) at a given age for female offspring in relation to their timing of birth (E, early; M, intermediate; L, late) for lowland (red) and highland (blue) populations. Estimates based on field data from 2000–2007 (details provided in the Supplementary Information). **b,** Annual variation in mean maximum temperature experienced during the first half of gestation for lowland (red) and highland (blue) populations.

female fitness is greatly reduced (Fig. 2a; Supplementary Table 3). Furthermore, highland populations experience relatively high between-year variance in temperature (Fig. 2b), which could select for GSD because it prevents extreme sex ratios and therefore reduces variance in fitness across breeding attempts[1,18–20].

To derive conditions under which the observed evolutionary divergence in sex determination in snow skinks could be favoured by natural selection, and to evaluate the relative importance of climate-induced changes in lizard life history and annual fluctuation in temperature, we constructed an individual-based simulation model based on a sex-determining mechanism recently proposed for lizards[5]. In this model, sex is determined by a threshold polymorphism involving four gene loci (see Supplementary Information for details). Each individual has a genetically determined temperature-dependent rate of regulatory gene expression, which needs to exceed a genetically determined threshold level to trigger male development (Supplementary Fig. 3). This allows evolutionary shifts in sex-determining systems via changes in the regulation of a developmental switch by genetic or environmental input. Both GSD and TSD can therefore be seen as emergent outcomes of selection for canalization of this switch, whereas 'mixed systems'[4,5] occur when canalization is incomplete (Supplementary Information). We parameterized this model with empirical data from long-term studies of two populations at the climatic extremes of the species' distribution and used sensitivity analyses to test whether climatic effects on life histories and the differences in the degree of between-year fluctuation in temperatures between altitudes were sufficient to explain the observed divergence in sex-determining systems. In addition, we calculated how well the temperatures experienced by individual females predicted their sex ratios to assess whether our model accurately captured the correlations observed in natural populations (see Methods and Supplementary Information for further details).

The model generated two primary results, both in close accordance with empirical data (Fig. 3). First, in simulations parameterized with data from the lowland population, sex determination evolved from pure GSD towards a system with a strong temperature effect (Fig. 3b). This
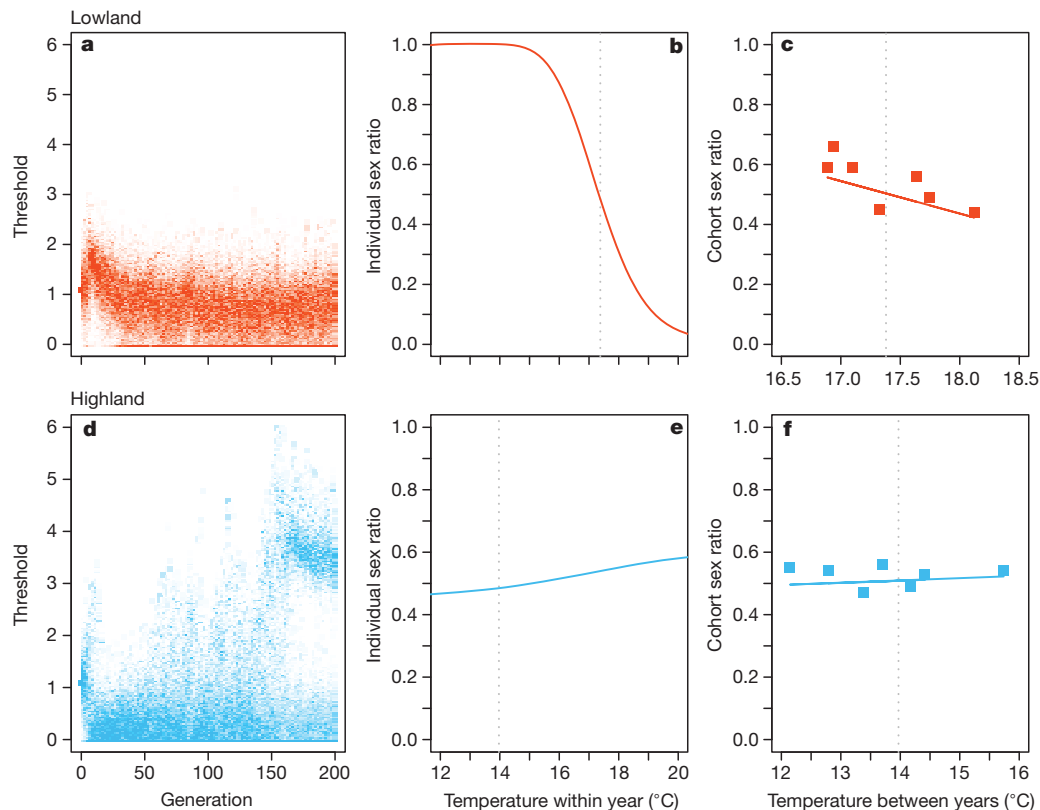
Figure 3 | Evolutionary simulation results with genetic sex determination as ancestral state. Upper panels, lowland parameter settings; lower panels, highland parameter settings. **a** and **d**, Population distributions of allelic values at threshold locus changing over time. We note branching in **d** for highland parameter settings, resulting in a novel sex-determining locus: males are 'homozygous' for alleles causing low thresholds and females 'heterozygous' for low and high threshold alleles. **b** and **e**, Evolved average reaction norm for offspring sex ratio as a function of developmental temperature. The vertical dotted line is the average temperature experienced by natural populations. **c** and **f**, Predicted (from evolved reaction norm; line) and observed (natural populations; squares) cohort sex ratios for annual mean maximum temperature in the wild.

generated a significant negative correlation between the cohort sex ratio and average temperature during gestation that closely resembled data from our natural population (Fig. 3c). Second, in simulations parameterized with data from the highland population, sex chromosomes (W or Y) of the initial GSD system were either retained or, if lost, were replaced by a novel genetic element of major effect via disruptive selection on the threshold locus (Fig. 3d). Consequently, the model could generate evolutionary shifts from one sex chromosome system to another—including transitions between male and female heterogamety (Supplementary Information)—but it always produced a sex-determining system that generated average sex ratios that did not deviate substantially from equality, again in close accordance with our natural population (Fig. 3e, f). These results were robust with respect to starting settings, male versus female heterogamety, and linkage between genetic elements (Supplementary Information).

The population divergence in sex-determining systems could be explained by both the increased rate of female maturation with earlier birth date in lowland population and the higher magnitude of annual fluctuations in temperature in the highland population (Supplementary Fig. 4). Thus, a relatively long activity season favours an evolutionary shift from GSD to TSD in lowland populations, manifested in our model through the loss of genes of major effect and adaptive evolution of a sex ratio reaction norm and hence TSD. Conversely, a relatively cold and more variable climate reduces the activity season and delays maturity, which results in minor birth date effects on female age and size at maturity and causes disruptive selection on regulatory elements in sex-determining networks and the emergence of novel sex chromosomes. This model may also capture observed population or species divergence in sex-determining systems in fish[10,12] and thus may be generally applied to short-lived species.

Climate-driven population divergence in sex-determining systems emphasizes a creative role of phenotypic plasticity in evolution[21]. First, the effect of climate on lizard life history is largely a passive result of how thermal opportunity constrains activity patterns rather than an evolved adaptation[22,23]. However, such non-adaptive plasticity can apparently contribute to divergent selection on seasonal sex ratio adjustment and, hence, sex-determining mechanisms across species' distributions. Second, the observation that stressfully high or low temperatures have a causal effect on sex determination also in vertebrates with GSD[5,24] suggests that temperature-induced developmental plasticity can simultaneously expose variation in sex determination and cause novel selection on this variation, thereby greatly facilitating evolutionary divergence in sex-determining systems[21,25]. If so, transitions between sex-determining systems may only require minor secondary modifications in the regulation of gonad differentiation, suggesting substantial scope for interchangeability between genetic and environmental determinants of sex[21].

METHODS SUMMARY

All data are based on field studies of two intensively monitored populations at the climatic extremes of the species' distribution[13,17,26] and from the Bureau of Meteorology station situated close to our study sites. Females undergo gestation in the field and are brought into the laboratory just before birth to enable assessment of sex ratios and reproductive output[26]. The data were used to estimate survival, onset of maturity and reproductive output as a function of birth date to generate parameter estimates for the simulation model (see Supplementary Information). We used the mean daily maximum temperatures during the period of temperature-sensitivity of embryos as our index of thermal opportunity[26,27].

To test directly the effect of thermal opportunity on sex determination we captured females early in gestation from areas adjacent to each of our main study sites and split them into two groups per population: extended basking conditions representative of warm years in lowland populations and limited basking conditions representative of cool years in highland populations (see ref. 14 for further detail).

Our simulation model is polygenic[18] and based on a dosage sex-determining mechanism recently proposed for lizards[5]. Sex is a threshold polymorphism determined by allelic values at four different loci (see Supplementary Information for details). We used daily temperatures from the past 20 years to calculate the long-term yearly mean (T_M) and the annual variation (σ_B) in temperature as well as the within-year variation (σ_W) in temperature. Each of 20 simulations started with 5,000 males and 5,000 females and the same values for reaction norm and threshold loci, with the age set to the minimum age at maturation. All results are from simulations run for 200,000 years.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

1. Bull, J. J. *The Evolution of Sex Determining Systems* (Benjamin/Cummings Inc., 1983).
2. Valenzuela, N. & Lance, V. *Temperature-Dependent Sex Determination in Vertebrates* (Smithsonian Books, 2004).
3. Janzen, F. J. & Phillips, P. C. Exploring the evolution of environmental sex determination, especially in reptiles. *J. Evol. Biol.* **19,** 1775–1784 (2006).
4. Radder, R., Pike, D. A., Quinn, A. E. & Shine, R. Offspring sex in a lizard depends on egg size. *Curr. Biol.* **19,** 1–4 (2009).
5. Quinn, A. E. *et al.* Temperature sex reversal implies sex gene dosage in a reptile. *Science* **316,** 411–416 (2007).
6. Crews, D. & Bull, J. J. Mode and tempo in environmental sex determination in vertebrates. *Semin. Cell Dev. Biol.* **20,** 251–255 (2009).
7. Warner, D. A. & Shine, R. The adaptive significance of temperature-dependent sex determination in a reptile. *Nature* **451,** 566–569 (2008).
8. Uller, T., Pen, I., Wapstra, E., Beukeboom, L. W. & Komdeur, J. The evolution of sex ratios and sex-determining systems. *Trends Ecol. Evol.* **22,** 292–297 (2007).
9. Charnov, E. & Bull, J. J. When is sex environmentally determined? *Nature* **266,** 828–830 (1977).
10. Conover, D. O. Adaptive significance of temperature-dependent sex determination in a fish. *Am. Nat.* **123,** 297–313 (1984).
11. Warner, D. A., Uller, T. & Shine, R. Fitness effects of the timing of hatching may drive the evolution of temperature-dependent sex determination in short-lived lizards. *Evol. Ecol.* **23,** 281–294 (2009).
12. Conover, D. O. & Heins, S. W. Adaptive variation in environmental and genetic sex determination in a fish. *Nature* **326,** 496–498 (1987).
13. Wapstra, E. & Swain, R. Geographic and annual variation in life history traits in a temperate zone Australian skink. *J. Herpetol.* **35,** 194–203 (2001).
14. Wapstra, E. *et al.* Maternal basking behaviour determines offspring sex in a viviparous reptile. *Proc. R. Soc. Lond. B* **271,** S230–S232 (2004).
15. Olsson, M. *et al.* Sexual dimorphism in lizard body shape: the roles of sexual selection and fecundity selection. *Evolution* **56,** 1538–1542 (2002).
16. Atkins, N. *Parental Effects in Two Species of Viviparous Lizards:* Niveoscincus microlepidotus *and* N. ocellatus. PhD thesis, University of Tasmania (2007).
17. Wapstra, E., Swain, R. & O'Reilly, J. M. Geographic variation in age and size at maturity in a small Australian viviparous skink. *Copeia* **3,** 646–655 (2001).
18. Bulmer, M. G. & Bull, J. J. Models of polygenic sex determination and sex ratio control. *Evolution* **36,** 13–26 (1982).
19. van Dooren, T. J. M. & Leimar, O. The evolution of environmental and genetic sex determination in fluctuating environments. *Evolution* **57,** 2667–2677 (2003).
20. Leimar, O., Hammerstein, P. & van Dooren, T. J. M. A new perspective on developmental plasticity and the principles of adaptive morph determination. *Am. Nat.* **167,** 367–376 (2006).
21. West-Eberhard, M.-J. *Developmental Plasticity and Evolution* (Oxford University Press, 2003).
22. Adolph, S. C. & Porter, W. P. Growth, seasonality, and lizard life histories: age and size at maturity. *Oikos* **77,** 267–278 (1996).
23. Cadby, C. D. *et al.* Multi-scale approach to understanding climate effects on offspring size at birth and date of birth in a reptile. *Integr. Zool.* **5,** 164–175 (2010).
24. Ospina-Alvarez, N. & Piferrer, F. Temperature-dependent sex determination in fish revisited: prevalence, a single sex ratio response pattern, and possible effects of climate change. *PLoS ONE* **3,** E2837 (2008).
25. Uller, T. & Badyaev, A. V. Evolution of "determinants" in sex determination: a novel hypothesis for the origin of environmental contingencies in avian sex-bias. *Semin. Cell Dev. Biol.* **20,** 304–312 (2009).
26. Wapstra, E. *et al.* Climate effects on offspring sex ratio in a viviparous lizard. *J. Anim. Ecol.* **78,** 84–90 (2009).
27. Neaves, L. *et al.* Embryonic gonadal and sexual organ development in a small viviparous skink, *Niveoscincus ocellatus. J. Exp. Zool.* **305A,** 74–82 (2006).

Author Contributions T.U., I.P. and E.W. initiated, planned and coordinated the project; E.W. collected field and experimental data, assisted by G.M.W.; T.U., G.M.W. and I.P. analysed data and generated parameter estimates for the model; I.P., B.F., A.H. and T.U. constructed the model and analysed its outcome; T.U. and I.P. wrote the paper with input from all other authors.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to I.P. (i.r.pen@rug.nl).

METHODS

Field procedures and data collection. Between 2000/2001 and 2007/2008 approximately 90% of females from one lowland and one highland population of *N. ocellatus* were captured every year at the end of gestation, just before giving birth, resulting in a total of >1,500 females and >4,500 offspring. The taxonomic status of the populations as a single species and details on differences in life history traits have been described elsewhere[13,17,28]. Females were housed in cages until parturition, when all offspring were measured and sexed using hemipene eversion (repeatability >0.98 on the basis of animals followed to sexual maturity)[26]. Sex in this species is determined during the first half of gestation[27]. Offspring were released back into their population of origin randomly at 12 locations within each population. Paternity was assessed in a subset of litters using microsatellites[16]. The field data was used to estimate survival, onset of maturity, and reproductive output as a function of birth date, which were subsequently used as parameter estimates for the simulation model (see below; Supplementary Table 1).

Common garden experiment. Females captured early in gestation (before sex determination is completed[27]) from areas adjacent to each of our main study sites were split into two groups per population: extended basking conditions representative of warm years in lowland populations (10 h of basking per 24 h) and limited basking conditions representative of cool years in highland populations (4 h of basking per 24 h)[13,14]. At parturition, offspring were measured and sexed as for the natural populations. Sex-specific mortality can be ruled out because the number of offspring corresponded to the number of ovulated eggs assessed using palpation.

Climate data. Climatic data was obtained from Bureau of Meteorology stations situated close to our study sites. As a measure of the thermal conditions (basking opportunity) experienced by individual female skinks while gravid in the field we used the mean of daily maximum temperatures during gestation (first half of gestation, assigned as 1 October to 15 November in lowland and 15 October to 1 December in highland populations), which is an accurate determinant of the temperature experienced during sex determination[26].

Simulation model. Our model is polygenic[18] and based on a dosage sex-determining system recently proposed for lizards[5]. Sex is a threshold polymorphism determined by allelic values at four different loci (see Supplementary Information for details). On the basis of daily temperatures from the past 20 years (from each altitude) we calculated the long-term yearly mean (T_M) and the annual variation (σ_B) in temperature as well as the within-year variation (σ_W) in temperature. In the model the yearly temperature (T_Y) is calculated at each time step by drawing a value from a normal distribution with mean T_M and standard deviation σ_B. T_Y is further used to calculate female-specific thermal conditions (T_F) by drawing a value from a normal distribution with mean T_Y and standard deviation σ_W. To facilitate model building, we divided each reproductive season into three categories: early, intermediate and late breeding (see Supplementary Information for further detail).

Data from our long-term study of two focal populations were used to estimate the minimum age at maturation, number of offspring, offspring and adult survival, and the probability of breeding at age t (Supplementary Information). Because age and body size do not influence male reproductive success in snow skinks[15,16], we set the effect of birth date on male reproductive fitness to be zero. Each of 20 simulations started with 5,000 males and 5,000 females and the same values for reaction norm and threshold loci, and with the age set to the minimum age at maturation. The life history follows a simple structure (Supplementary Fig. 1). In brief, females mate with a randomly drawn male and produce a number of offspring according to her age drawn from a distribution of clutch sizes. The sex of the offspring is determined by the number of Z (or X) chromosomes, the reaction norm and threshold loci, and T_F (Supplementary Fig. 3). Offspring have a fixed probability of survival to the next year (survival is independent of birth date; Supplementary Information). Offspring that have reached the minimum age at maturation have a fixed age-specific probability of reproducing that depends on their timing of birth. At the end of each time step all individuals in the population age by one year and the cycle is restarted. All results are from simulations run for 200,000 years.

28. Melville, J. & Swain, R. Evolutionary relationships between morphology, performance and habitat openness in the lizard genus *Niveoscincus* (Scincidae: Lyosomaniae). *Biol. J. Linn. Soc.* **70,** 667–680 (2000).

ARTICLE

Support for a synaptic chain model of neuronal sequence generation

Michael A. Long[1]†, Dezhe Z. Jin[2] & Michale S. Fee[1]

In songbirds, the remarkable temporal precision of song is generated by a sparse sequence of bursts in the premotor nucleus HVC. To distinguish between two possible classes of models of neural sequence generation, we carried out intracellular recordings of HVC neurons in singing zebra finches (*Taeniopygia guttata*). We found that the subthreshold membrane potential is characterized by a large, rapid depolarization 5–10 ms before burst onset, consistent with a synaptically connected chain of neurons in HVC. We found no evidence for the slow membrane potential modulation predicted by models in which burst timing is controlled by subthreshold dynamics. Furthermore, bursts ride on an underlying depolarization of ~10-ms duration, probably the result of a regenerative calcium spike within HVC neurons that could facilitate the propagation of activity through a chain network with high temporal precision. Our results provide insight into the fundamental mechanisms by which neural circuits can generate complex sequential behaviours.

Complex behaviours are made possible by the ability of the brain to step through well defined sequences of neural states[1]. Brain processes capable of generating intrinsic sequential activity are thought to underlie motor sequencing[2], navigation[3,4], movement planning[5], sensitivity to the timing of sensory stimuli[6] and cognitive tasks[7]. With few exceptions[8], however, the biophysical mechanisms by which neural circuits produce sequences are poorly understood.

Songbirds have emerged as an excellent model system for investigating the neural mechanisms of sequence generation. The adult zebra finch song motif consists of a stereotyped pattern of song syllables[9]. One premotor forebrain area in particular, nucleus HVC (used as a proper name), is known to have a central role in controlling the temporal structure of birdsong[10–12]. During singing, neurons in HVC projecting to downstream premotor nucleus RA (robust nucleus of the arcopallium) produce only a single highly stereotyped burst of spikes during each repetition of the song motif[13]. Different RA-projecting HVC neurons (HVC$_{(RA)}$) burst at different time points in the song, indicating that HVC neurons may burst sequentially through the song motif, in turn activating a complex and highly stereotyped pattern of bursts in the downstream nucleus RA[14,15].

Here we set out to distinguish experimentally among several distinct classes of possible sequence-generating circuits within HVC. First, it has been proposed that sequential states of neural activity may be generated by synaptically connected chains of neurons[6,16,17]. In this view, activity could propagate through the HVC network—like a chain of falling dominoes—forming the basic clock that underlies song timing (Fig. 1a)[10,18–20]. A second, fundamentally different, class of models can allow for sequence generation in the absence of overt feed-forward connections between HVC$_{(RA)}$ neurons. In these models, oscillatory or other subthreshold dynamics can modulate the excitability of neurons and thus control the timing of their activity[21,22], like those proposed to control the sequential activation of spikes during hippocampal theta sequences[23] and within replay events[3,24]. Subthreshold dynamics and rhythmicity on the timescale of song syllables (~100 ms) exist within HVC *in vitro*[25] and thus could have a central

role in controlling the timing of HVC bursts on that timescale in the singing bird.
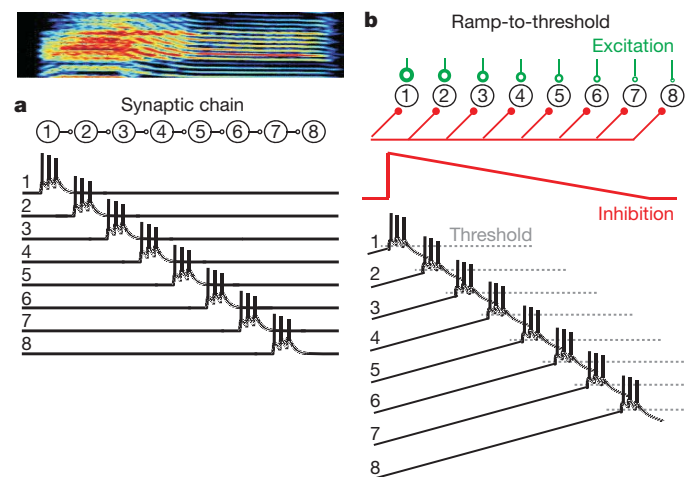
Figure 1 | Two broad classes of models for a sequence-generating circuit. a, Neurons might form a feed-forward synaptically connected chain within the HVC such that activity propagates from one group of neurons to the next. **b,** Alternatively, sequential activity might occur in the absence of directed connections between neurons, from temporal and spatial gradients of excitability. For example, the network could receive a global and gradual ramping-down of an inhibitory input over time (red synapses), producing a sequential activation. The order of activation would be determined by neuronal excitability. In the example model shown here, neurons receive different levels of constant excitatory input (green synapses). The neuron with the largest excitatory input (neuron 1) would be most depolarized and would be the first to reach spiking threshold. The neuron with the smallest constant excitatory input (neuron 8) would be the last to reach threshold. In the model depicted here, the timescale of the sequence produced corresponds to one song syllable (shown above).

[1]McGovern Institute for Brain Research, Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, Massachusetts 02139, USA. [2]Department of Physics, The Pennsylvania State University, University Park, Pennsylvania 16802, USA. †Present address: Departments of Otolaryngology and Physiology and Neuroscience, NYU School of Medicine, 522 First Avenue, New York, New York 10016, USA.

Intracellular recording during singing

To examine the role of subthreshold dynamics in the control of timing of HVC bursts during singing, we adopted an approach recently introduced for intracellular recordings in the freely moving rat[26]. We developed a miniature (1.6 g) microdrive that allows sharp microelectrode recordings to be performed in singing male zebra finches (Fig. 2a). Birds could move freely in a recording chamber, unrestrained except for a thin, flexible tether. In total, 28 neurons in 12 birds were recorded during singing of all three HVC neuron types, defined broadly by their axonal projections[27,28] (Fig. 2b).

The singing-related spiking patterns of intracellularly recorded neurons closely resembled the previously described patterns in extracellular recordings[13,29]. Putative interneurons ($n = 3$) were identified by a high spontaneous firing rate, and a continuous high firing rate throughout song (Fig. 2c, 117 ± 24.6 Hz singing, 66.3 ± 21.6 Hz baseline, error bars indicate $\pm$s.e.m. unless otherwise noted). Putative HVC neurons projecting to the basal ganglia homologue area X ($n = 12$) exhibited a low spontaneous spiking rate (<10 Hz) when the bird was not singing, and one or more high-frequency bursts during singing (Fig. 2d). These neurons showed a gradual hyperpolarization during the introductory notes (before the first motif in a bout of singing), and were hyperpolarized during song motifs (Fig. 2d, $n = 12$ of 12 cells; singing, -70.8 ± 3.4 mV; baseline, -67.7 ± 3.1 mV), similar to what has been observed during auditory song playback[30,31]. We did not consider these neurons further in the context of sequence generation because it has been shown that selective ablation of X-projecting HVC neurons in adult zebra finches does not impair song production[32].

HVC neurons that project to RA were identified by antidromic stimulation from RA (Fig. 2e, inset; see also Supplementary Fig. 1)[13]. HVC$_{(RA)}$ neurons showed a gradual depolarization before the onset of singing (Fig. 2e) and were persistently depolarized during singing ($n = 13$ of 13 cells; singing, -67.3 ± 3.5 mV; baseline, -75.7 ± 3.5 mV). About half of HVC$_{(RA)}$ neurons ($n = 7$ of 13) generated a single burst during each song motif (Fig. 3a–c, 3.8 ± 0.6 spikes per burst). The remaining HVC$_{(RA)}$ neurons ($n = 6$ of 13 cells) did not spike during song motifs (for example, Fig. 3d)[13].

Chain model versus ramp-to-threshold model

Recurrent synaptic connections within a network of sequentially active neurons would be expected to produce patterned synaptic inputs; thus previous reports of patterned synaptic inputs have been used as evidence of synaptically connected chains both *in vitro* and *in vivo*[33]. Consistent with this view, we observed a highly stereotyped pattern of fast subthreshold fluctuations widely distributed throughout the song (Figs 2e and 3a–d, and Supplementary Fig. 2). For individual neurons, the song-aligned subthreshold fluctuations were highly correlated across song motifs (cross-correlation 0.80 ± 0.04, $P < 10^{-9}$, $n = 13$ neurons).

We now ask whether, as predicted by the ramp-to-threshold model, there was any slow ramping of membrane potential before the onset of bursts (Fig. 1b). We first consider the time window from the beginning of the song motif to the burst onset for each neuron. Across all 7 HVC$_{(RA)}$ neurons that burst during singing, the membrane potential did not change significantly in the period from the beginning of the
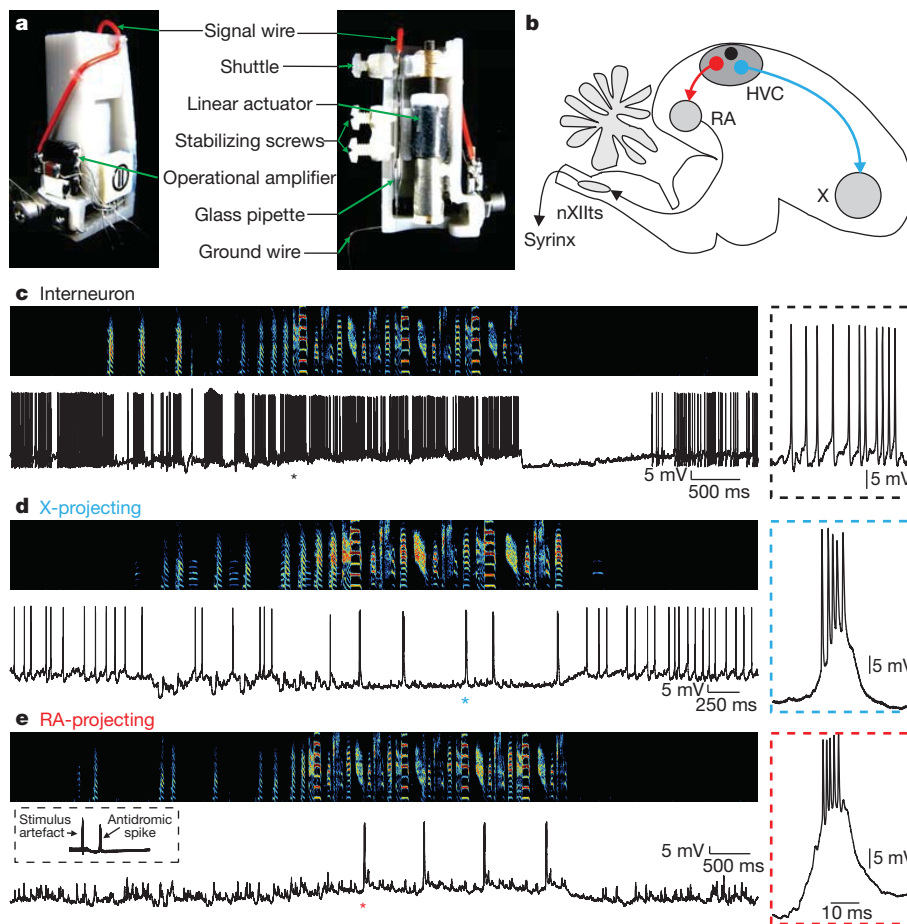
Figure 2 | A microdrive for sharp intracellular recording in the singing bird. **a**, The intracellular microdrive incorporates a motor that rotates a threaded rod and advances a shuttle that holds the electrode. **b**, A schematic of the zebra finch brain, highlighting three cell types in nucleus HVC defined by their projections: local circuit interneurons (in black), neurons that project to RA (in red), and neurons that project to basal-ganglia-homologue area X (in blue). **c–e**, Examples of intracellular records from a putative local circuit interneuron (**c**), a putative X-projecting neuron (**d**) and an antidromically identified RA-projecting neuron (**e**). Asterisk indicates the region magnified in the panels to the right.
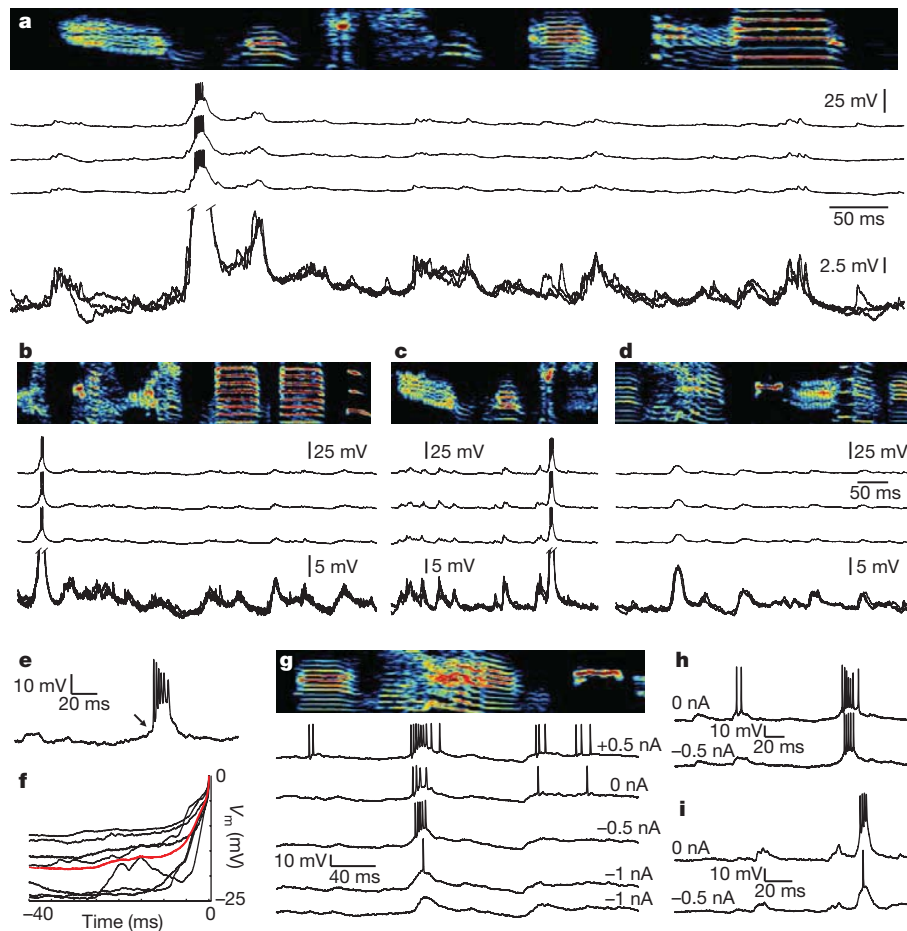
Figure 3 | Intracellular membrane potential of identified HVC$_{(RA)}$ neurons during singing. **a–d**, Examples of the membrane potential of four HVC$_{(RA)}$ neurons recorded during singing. For each cell, activity from three motif renditions is shown aligned to the song (top). Also shown is an overlay of the membrane potential traces (expanded vertical scale, bottom of each panel). **e**, Expanded view of a burst from another neuron during singing showing the flat membrane potential before burst onset (arrow). **f**, Average membrane potential of seven HVC$_{(RA)}$ neurons before the first spike in the burst (time zero). The population average is shown in red. **g–i**, The membrane potential of three HVC$_{(RA)}$ neurons during singing with different holding currents. **g**, One neuron was held long enough to record with injected currents of +0.5 nA, 0 nA, −0.5 nA and −1.0 nA. **h, i**, Two other neurons recorded with 0 nA and −0.5 nA hyperpolarizing current. Note that injected current had little effect on burst timing, inconsistent with the predictions of the ramp-to-threshold model.

song motif to the moment 10 ms before the first spike in the burst (−0.47 ± 0.69 mV, $P = 0.53$, t-test, average window duration, 387 ± 92 ms). We next considered a ramp of excitation on the shorter timescale of a song syllable (~100 ms). Across all bursting neurons ($n = 7$), the membrane potential did not change during a window from 100 ms to 10 ms before the first spike in the burst (0.31 ± 1.04 mV, $P = 0.77$, t-test). Both of these results are inconsistent with a slow ramp of excitation before burst onset, on the timescale of either a song motif or a song syllable. In contrast, bursts of HVC$_{(RA)}$ neurons were preceded, within the 5 ms before the first spike in the burst, by a large depolarization of 10.5 ± 1.9 mV from baseline (Fig. 3e, f, the first spike of the burst initiated at a membrane potential of −52.6 ± 1.7 mV). This result is consistent with a model in which HVC$_{(RA)}$ neurons are activated by a large synchronous synaptic input from a group of previously active neurons.

The two models described in Fig. 1 give very different predictions for the effect of intracellular current injection on the timing of neural activity. In a model in which the timing of HVC$_{(RA)}$ bursts is controlled by slow membrane potential dynamics (Fig. 1b), an injected depolarizing current would cause the neuron to burst earlier during the slow depolarizing ramp, assuming that the burst-generating mechanism is sufficiently well coupled to the site of current injection (see Supplementary Discussion). In contrast, in the chain model, burst timing is controlled by a synaptic input from a preceding group of neurons (Fig. 1a). Thus, current injection would have a minimal

effect on burst timing, perhaps causing the first spike in the burst to appear a few milliseconds earlier during the onset of the synaptic depolarization.

We assessed the effect of intracellular current injection on the timing of bursts in HVC$_{(RA)}$ neurons during singing in three neurons. Two neurons were recorded with zero holding current and with 0.5 nA of hyperpolarizing current. One additional neuron was held long enough to record at four levels of holding current (0.5 nA, 0 nA, −0.5 nA and −1 nA). On average, the resulting membrane potential change was 20.3 mV nA^{-1} of injected current. In all cases, hyperpolarizing current was seen to reduce the number of spikes in the burst (Fig. 3g–i, average 5 spikes per burst at 0 nA compared to 3.3 spikes per burst at −0.5 nA), and could suppress spiking completely at the most hyperpolarizing currents (−1.0 nA). Depolarizing current injection increased the number of spikes per burst (Fig. 3g).

Remarkably, the timing of the burst was only weakly affected by injected currents. At a hyperpolarizing holding current of 0.5 nA, the burst onset was delayed by an average of only 2.6 ms ($n = 3$). However, the last spike of the burst was advanced by a similar amount such that the centre of the burst (midpoint between first and last spikes) was very weakly affected by injected current (1.2 ms nA^{-1}, Fig. 3g–i). In addition, under conditions at which the spiking was suppressed or nearly suppressed by hyperpolarizing current, a large underlying depolarization at the temporal position of the burst was clearly visible (Fig. 3g, i). These results are consistent with a mechanism in which a given

HVC(RA) neuron is driven by fast synaptic input from a preceding group of neurons.

Cellular mechanisms of burst generation

The broad powerful depolarizations that underlie the bursts of spikes in HVC(RA) neurons during singing (Fig. 3) are reminiscent of dendritic calcium spikes observed in many neurons[34,35]. Although it is difficult to establish definitively that the singing-related bursts of HVC(RA) neurons are mediated by calcium spikes, we have carried out *in vitro* and *in vivo* whole-cell recordings and pharmacological manipulations that support this view.

Although HVC(RA) neurons have not been observed to generate a burst response to somatic intracellular current injection (Fig. 4a, b)[27,28,36], dendritic calcium spikes in some neurons may not be observed during somatic current injection[37], but can be unmasked by the intracellular blockade of sodium and potassium channels[38]. We carried out whole-cell recordings in brain slices of antidromically identified HVC(RA) neurons with QX-314 in the recording pipette. Indeed, current injection resulted in a large depolarizing event in all neurons tested ($n = 23$ cells, Fig. 4c, average amplitude 26.4 ± 5.6 mV, width at half height 4.5 ± 1.0 ms). The depolarizing events had a clear all-or-none response with an initiation threshold at the soma of -36.2 ± 4.4 mV (Fig. 4d, $n = 14$ cells, compared to a threshold of -40.3 ± 4.3 mV for sodium spikes). In contrast, neurons in nucleus RA did not exhibit all-or-none spikes in the presence of QX-314 (ref. 39; Supplementary Fig. 3). The depolarizing events in HVC(RA) neurons were completely blocked by the broad spectrum calcium channel antagonist cadmium (100 μM, $n = 4$ cells), but were unaffected by nickel (100 μM, $n = 5$ cells), an antagonist of low-threshold voltage-gated calcium channels, indicating that the depolarizing events might be mediated by a high-threshold calcium channel.

We found that the L-type calcium channel agonist BAY K 8644 could enhance the calcium current sufficiently to evoke a burst response in HVC(RA) neurons even in the absence of QX-314 ($n = 8$ cells, Fig. 4e, f, average of 3.4 ± 0.2 spikes, within-burst spike rate 302 ± 14 Hz). These burst responses appeared to have an all-or-none characteristic with a well-defined threshold for injected current (0.50 ± 0.05 nA), and a spike rate within bursts that did not increase at higher currents ($P = 0.60$). These *in vitro* experiments indicate that HVC(RA) neurons are capable, under some conditions, of generating calcium-based regenerative spikes, possibly mediated by an L-type Ca conductance.

We wanted to examine more directly the role of these calcium conductances under conditions in which HVC(RA) neurons naturally generate burst sequences, rather than in brain slice. In a form of 'replay' of song-like patterns[40], HVC(RA) and RA neurons generate sparse sequential bursts during sleep similar to those produced during singing[13,41]. We have adapted a head-fixed sleeping bird preparation[13] and used whole-cell recordings and pharmacological manipulation of HVC(RA) neurons to study the mechanisms underlying these bursts in naturally sleeping zebra finches (Fig. 4g). Across the population of HVC(RA) neurons in our data set ($n = 36$ cells), nearly half the spikes recorded ($49.3 \pm 3.5\%$) formed high-frequency bursts (>100 Hz) during sleep (2.74 ± 0.11 sodium spikes, average within-burst rate of 265 ± 13 Hz). Just as during singing, sleep bursts were seen to ride on a prominent underlying depolarizing event (Fig. 4g, 25.2 ± 0.9 mV amplitude, 18.4 ± 1.5 ms width at 2/3 height).

Injections of the L-type calcium channel agonist BAY K 8644 (100 μM, 5–20 nl bolus) in the vicinity (<100 μm) of the whole-cell recording pipette increased the burst size (Fig. 4i, increased number of spikes and total burst duration, $P < 10^{-5}$ for both measures, Kolmogorov-Smirnov test). In addition, these injections significantly increased the incidence of bursting (Fig. 4k, mean interburst interval 2.0 ± 5.7 s with BAY K 8644, compared to 18.4 ± 34.5 s control, $P < 10^{-4}$, Kolmogorov-Smirnov test, $n = 6$ cells from 5 birds, mean ± s.d.). In contrast, injections of the L-type calcium channel antagonist nifedipine (100 μM) significantly decreased burst incidence (Fig. 4k,
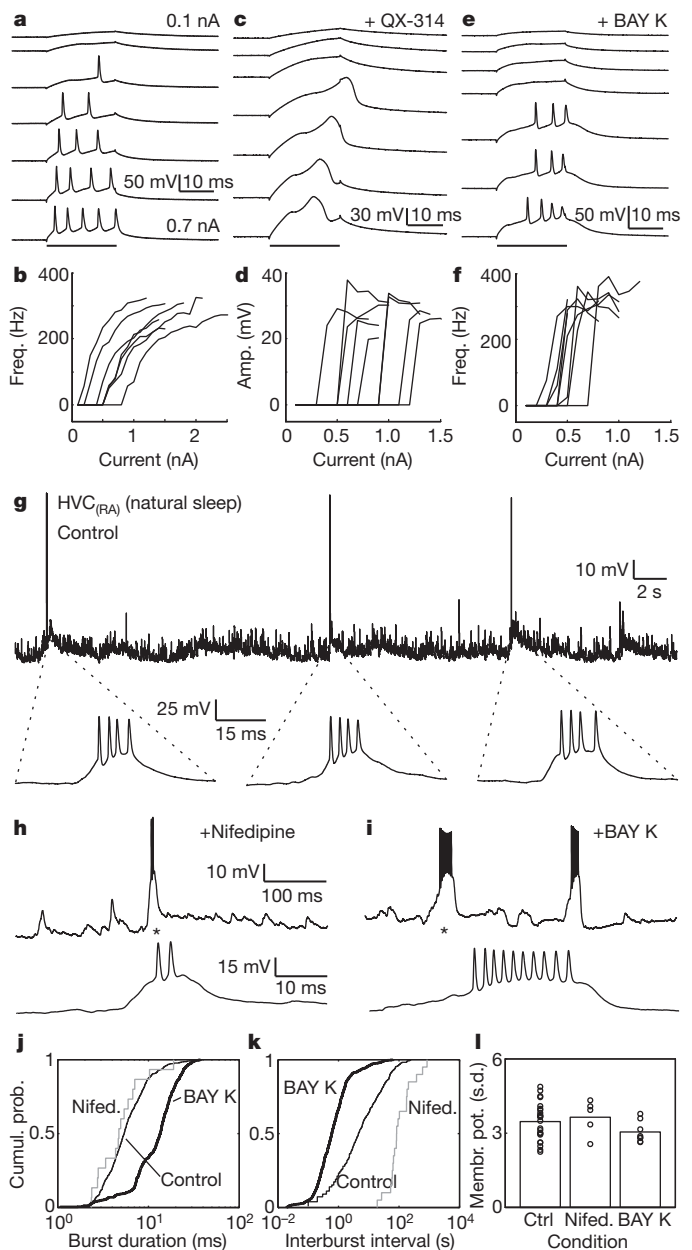
Figure 4 | Evidence that calcium channels contribute to burst events in HVC(RA) neurons. **a**, Response of an HVC(RA) neuron in brain slice to somatically injected current steps (black bar) of different size. **b**, Relationship between injected current and evoked firing rate in a population of 7 HVC(RA) neurons. Note that somatic current injection does not elicit an all-or-none burst. **c**, In the presence of intracellular sodium and potassium channel blocker QX-314 (5 mM), calcium spikes appear as an all-or-none depolarizing event. **d**, The amplitude of the depolarizing event (threshold to maximum point) as a function of injected current reveals an all-or-none response ($n = 8$ of 8 cells). **e, f**, HVC(RA) neurons treated with the L-type calcium channel agonist BAY K 8644 (5–10 μM) generate all-or-none spike bursts in response to somatic current injection. **g**, Segment of a whole-cell recording in a head-fixed bird during natural sleep showing three spontaneous bursts. **h, i**, Spontaneous bursting activity recorded during sleep after localized injection of L-type calcium channel antagonist nifedipine (**h**) or agonist BAY K 8644 (**i**). Asterisk indicates expanded view below. **j, k**, Cumulative distribution of burst durations and inter-burst intervals for control, nifedipine and BAY K 8644 conditions. **l**, Standard deviation of membrane potential fluctuations is not affected by nifedipine or Bay K 8644, indicating that synaptic transmission is not affected by these drugs.

mean interburst interval 171.7 ± 209.6 s, greater than control, $P < 0.0001$, Kolmogorov-Smirnov test, $n = 6$ cells from 4 birds, mean ± s.d.). The effect of L-type calcium channel modulators could

not be explained by changes in the size of synaptic inputs: the magnitude of fluctuations in membrane potential was not altered by BAY K 8644 or nifedipine ($P > 0.05$, t-test, Fig. 4l). Taken together, these experiments demonstrate that L-type calcium channels have a role in generating or initiating bursting activity in $HVC_{(RA)}$ neurons. Such highly nonlinear all-or-none calcium spikes produce a highly stereotyped response to a wide range of synaptic inputs[42], and could have implications for the propagation of activity in a synaptically connected chain of neurons.

Burst propagation in a chain network

The stable propagation of bursts in an excitatory chain network is non-trivial; it requires precisely tuned synaptic strengths to avoid runaway excitation or decay[19]. It has previously been shown that an intrinsic neuronal burst mechanism can allow the stable propagation of activity in a chain network[19], but what about temporal precision and stereotypy? Here we use a simple biophysical model to examine the role that intrinsic bursting might have in achieving precise stereotyped temporal structure in the presence of noise. We also examine how such a mechanism might make the functioning of these networks robust over a wide range of network and synaptic properties.

We studied a network of 70 groups of 30 excitatory $HVC_{(RA)}$ neurons each, organized in a sequentially connected chain. Recurrent inhibition in HVC[25,43] was implemented by a population of 300 interneurons with sparse random connections to the excitatory chain (Supplementary Fig. 4a). We began with a non-bursting model of $HVC_{(RA)}$ neurons, described by a single spiking somatic compartment (Fig. 5, Supplementary Fig. 4b and Supplementary Methods).
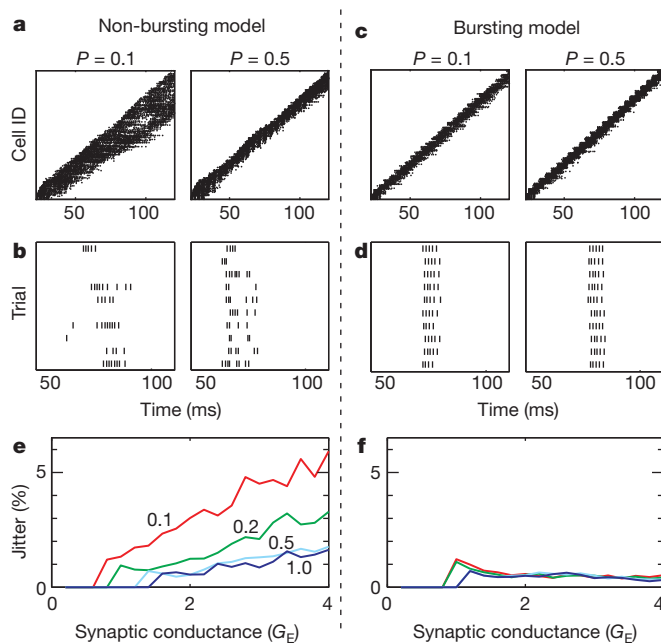
Figure 5 | A simple biophysical model to examine the implications of neuronal bursting on the robustness of HVC network propagation.
a–d, Two models of a synaptically connected chain network were compared: one with non-bursting neurons (**a**, **b**), the other with bursting neurons (**c**, **d**). **a**, Non-bursting model: spike raster plot for all neurons in the network showing activity as a function of time for two different levels of network connection probability ($P = 0.1$ and 0.5). **b**, Spike raster of a single neuron during different runs of the network. Note the non-stationarity of propagation and large variability across runs. **c**, Bursting model: spike raster plot for all neurons in the network. **d**, Spike raster of a single neuron during different runs of the network. Note the highly uniform propagation and stereotyped response across runs. **e**, **f**, Run-time jitter, plotted as a function of network connectivity and synaptic conductance, is consistently lower in the bursting model than in the non-bursting model. (See Supplementary Figures and Table for further quantification, and Supplementary Methods for model details.)

We found that this network did not exhibit the unstable (explosive or decaying) behaviour characteristic of purely excitatory networks[19], but exhibited stable propagation of burst activity over a wide range of connection probabilities ($P = 0.1$–1.0) and excitatory synaptic strengths between $HVC_{(RA)}$ neurons in successive groups ($G_{EE,max}$ from 0.2 to 4.0 mS cm^{-2}). Nevertheless, the activity tended to be non-stationary, particularly at lower connection probabilities ($P = 0.1$, Fig. 5a), exhibiting both dispersion (broadening) and variations in propagation velocity at different points in the network (Supplementary Figs 5 and 6a, b). Furthermore, the network was sensitive to the presence of noise, producing activity that was not stereotyped across multiple trials of the simulation, including large jitter in the speed of propagation through the network (Fig. 5b, e; $1.95 \pm 1.38\%$ mean run-time jitter $\pm$ s.d.) and large variations in the burst response on different trials (quantified as spikes per burst and burst unreliability, Supplementary Fig. 6c–e). Finally, many characteristics of the propagation (number of spikes per burst, burst duration and burst jitter) were strongly dependent on the network connection probabilities and connection strengths (Fig. 5 and Supplementary Fig. 6). Thus, although the stable propagation of bursts is possible in a chain network of non-bursting neurons, the network does not produce the stereotyped sequences characteristic of real $HVC_{(RA)}$ neurons.

The situation was markedly different in a model with neurons that have an intrinsic burst mechanism. Bursting $HVC_{(RA)}$ neurons were modelled with a spiking somatic compartment plus a dendritic compartment containing conductances for generating calcium spikes (see Supplementary Figs 4c, d and 7–9). Propagation down the chain was stationary, with no broadening or variations in velocity (Fig. 5c and Supplementary Fig. 5). The propagation was also extremely stereotyped, exhibiting small trial-to-trial variations in propagation speed (Fig. 5d, $0.52 \pm 0.17\%$ mean run-time jitter $\pm$ s.d.). Burst response was much more reliable in the bursting model (see spikes per burst and burst unreliability, Supplementary Fig. 6), similar to what has been observed in singing-related firing patterns of $HVC_{(RA)}$ neurons[13,29]. Finally, in the bursting model, every characteristic of burst propagation that we examined was much more robust to variations in network connection probability and synaptic strength than was the single compartment model (Fig. 5e, f and Supplementary Fig. 6). Similar results were obtained with a simple integrate-and-burst model (Supplementary Fig. 10). Taken together, these results indicate that an intrinsic neuronal burst mechanism, regardless of its biophysical implementation, could serve a fundamental role in allowing synaptically connected chain networks to propagate in a highly stereotyped manner with low temporal jitter, even in the presence of noise, and over a wide range of network connectivities. Such robustness could also make sequence-generating networks easier to assemble during development[44,45].

We have carried out intracellular recording and manipulation of activity in the freely behaving animal in a neural circuit important for the temporal control of behaviour. We observed no ramping or rhythmicity that could contribute to the temporal patterning of $HVC_{(RA)}$ bursts. In contrast, our recordings reveal a single large postsynaptic potential that immediately precedes the onset of a song-locked burst of spikes. Together, our findings are consistent with the idea that the control of song temporal structure is produced by the propagation of calcium-mediated bursts through a synaptically connected chain of neurons. Temporally precise learned behaviours in other vertebrates could use similar mechanisms to organize neuronal activity into sequentially active states.

METHODS SUMMARY

Subjects. We used adult (>120 days post hatch) male zebra finches (*Taeniopygia guttata*). All animal procedures were reviewed and approved by the MIT committee on animal care.
Intracellular recording during singing. Intracellular recordings were achieved in the zebra finch using a custom microdrive constructed out of 3D printed plastic (AP Proto) outfitted with a lightweight linear actuator (Smoovy Series 0515,

Faulhaber). A preamplifer was mounted at the base of the device which routed signals to a commercially available intracellular amplifier (IR-183, Cygnus Technology). Sharp microelectrodes were pulled to a final impedance of 80–110 MΩ and were filled with 3 M potassium acetate. Once a stable intracellular recording was obtained, a female bird was presented to elicit directed singing.

Intracellular recording during sleep. During an initial surgical step, a stainless steel headplate was affixed to the skull. A small ($\sim$200 µm) craniotomy was made over HVC. Whole-cell recordings were made with glass electrodes (5–8 MΩ) using techniques described elsewhere[46]. Signals were measured using an Axoclamp 2B (Molecular Devices). In some experiments, an injection pipette (20–30 µm opening) was positioned less than 100 µm from the recording site for the injection (Nanoject II, Drummond Scientific) of a small volume (5–20 nl) of 100 µM ($+/-$)-BAY K 8644 (A.G. Scientific) or 100 µM nifedipine (Sigma).

Slice preparation. 400-µm slices were prepared on a vibrating microtome (Leica VT1000) and placed in ice-cold ACSF (sodium replaced with equimolar sucrose). Slices were then recorded in an interface-style chamber (VB5000, Leica) with standard ACSF (in mM): 126 NaCl, 3 KCl, 1.25 NaH_2PO_4, 2 $MgSO_4$-7H_2O, 26 $NAHCO_3$, 10 dextrose, 2 $CaCl_2$-2$H_2$0. QX-314 (5 mM, internal) was used in a subset of these experiments.

1. Lashley, K. in *Cerebral Mechanisms in Behavior* (ed. Jeffress, L.) (Wiley, 1951).
2. Shima, K., Isoda, M., Mushiake, H. & Tanji, J. Categorization of behavioural sequences in the prefrontal cortex. *Nature* **445,** 315–318 (2007).
3. Mehta, M. R., Lee, A. K. & Wilson, M. A. Role of experience and oscillations in transforming a rate code into a temporal code. *Nature* **417,** 741–746 (2002).
4. Harvey, C. D., Collman, F., Dombeck, D. A. & Tank, D. W. Intracellular dynamics of hippocampal place cells during virtual navigation. *Nature* **461,** 941–946 (2009).
5. Pastalkova, E., Itskov, V., Amarasingham, A. & Buzsaki, G. Internally generated cell assembly sequences in the rat hippocampus. *Science* **321,** 1322–1327 (2008).
6. Mauk, M. D. & Buonomano, D. V. The neural basis of temporal processing. *Annu. Rev. Neurosci.* **27,** 307–340 (2004).
7. Georgopoulos, A. P., Lurito, J. T., Petrides, M., Schwartz, A. B. & Massey, J. T. Mental rotation of the neuronal population vector. *Science* **243,** 234–236 (1989).
8. Stent, G. S. *et al.* Neuronal generation of the leech swimming movement. *Science* **200,** 1348–1357 (1978).
9. Konishi, M. Birdsong: from behavior to neuron. *Annu. Rev. Neurosci.* **8,** 125–170 (1985).
10. Long, M. A. & Fee, M. S. Using temperature to analyse temporal dynamics in the songbird motor pathway. *Nature* **456,** 189–194 (2008).
11. Yu, A. C. & Margoliash, D. Temporal hierarchical control of singing in birds. *Science* **273,** 1871–1875 (1996).
12. Vu, E. T., Mazurek, M. E. & Kuo, Y. C. Identification of a forebrain motor programming network for the learned song of zebra finches. *J. Neurosci.* **14,** 6924–6934 (1994).
13. Hahnloser, R. H., Kozhevnikov, A. A. & Fee, M. S. An ultra-sparse code underlies the generation of neural sequences in a songbird. *Nature* **419,** 65–70 (2002).
14. Leonardo, A. & Fee, M. S. Ensemble coding of vocal control in birdsong. *J. Neurosci.* **25,** 652–661 (2005).
15. Fee, M. S., Kozhevnikov, A. A. & Hahnloser, R. H. Neural mechanisms of vocal sequence generation in the songbird. *Ann. NY Acad. Sci.* **1016,** 153–170 (2004).
16. Amari, S. Learning patterns and pattern sequences by self-organizing nets of threshold elements. *IEEE Trans. Comp.* **c-21,** 1197–1206 (1972).
17. Abeles, M. *Corticonics: Neural Circuits of the Cerebral Cortex* (Cambridge Univ. Press, 1991).
18. Li, M. & Greenside, H. Stable propagation of a burst through a one-dimensional homogeneous excitatory chain model of songbird nucleus HVC. *Phys. Rev. E* **74,** 011918 (2006).
19. Jin, D. Z., Ramazanoglu, F. M. & Seung, H. S. Intrinsic bursting enhances the robustness of a neural network model of sequence generation by avian brain area HVC. *J. Comput. Neurosci.* **23,** 283–299 (2007).
20. Glaze, C. M. & Troyer, T. W. Behavioral measurements of a temporally precise motor code for birdsong. *J. Neurosci.* **27,** 7631–7639 (2007).
21. Mann, E. O. & Paulsen, O. Role of GABAergic inhibition in hippocampal network oscillations. *Trends Neurosci.* **30,** 343–349 (2007).
22. O'Keefe, J. & Recce, M. L. Phase relationship between hippocampal place units and the EEG theta rhythm. *Hippocampus* **3,** 317–330 (1993).
23. Foster, D. J. & Wilson, M. A. Hippocampal theta sequences. *Hippocampus* **17,** 1093–1099 (2007).
24. Buzsaki, G. Two-stage model of memory trace formation: a role for ''noisy'' brain states. *Neuroscience* **31,** 551–570 (1989).
25. Solis, M. M. & Perkel, D. J. Rhythmic activity in a forebrain vocal control nucleus *in vitro*. *J. Neurosci.* **25,** 2811–2822 (2005).
26. Lee, A. K., Manns, I. D., Sakmann, B. & Brecht, M. Whole-cell recordings in freely moving rats. *Neuron* **51,** 399–407 (2006).
27. Dutar, P., Vu, H. M. & Perkel, D. J. Multiple cell types distinguished by physiological, pharmacological, and anatomic properties in nucleus HVc of the adult zebra finch. *J. Neurophysiol.* **80,** 1828–1838 (1998).
28. Mooney, R. Different subthreshold mechanisms underlie song selectivity in identified HVc neurons of the zebra finch. *J. Neurosci.* **20,** 5420–5436 (2000).
29. Kozhevnikov, A. A. & Fee, M. S. Singing-related activity of identified HVC neurons in the zebra finch. *J. Neurophysiol.* **97,** 4271–4283 (2007).
30. Lewicki, M. S. Intracellular characterization of song-specific neurons in the zebra finch auditory forebrain. *J. Neurosci.* **16,** 5855–5863 (1996).
31. Rosen, M. J. & Mooney, R. Inhibitory and excitatory mechanisms underlying auditory responses to learned vocalizations in the songbird nucleus HVC. *Neuron* **39,** 177–194 (2003).
32. Scharff, C., Kirn, J. R., Grossman, M., Macklis, J. D. & Nottebohm, F. Targeted neuronal death affects neuronal replacement and vocal behavior in adult songbirds. *Neuron* **25,** 481–492 (2000).
33. Ikegaya, Y. *et al.* Synfire chains and cortical songs: temporal modules of cortical activity. *Science* **304,** 559–564 (2004).
34. London, M. & Hausser, M. Dendritic computation. *Annu. Rev. Neurosci.* **28,** 503–532 (2005).
35. Murayama, M. *et al.* Dendritic encoding of sensory stimuli controlled by deep cortical interneurons. *Nature* **457,** 1137–1141 (2009).
36. Kubota, M. & Taniguchi, I. Electrophysiological characteristics of classes of neuron in the HVc of the zebra finch. *J. Neurophysiol.* **80,** 914–923 (1998).
37. Larkum, M. E. & Zhu, J. J. Signaling of layer 1 and whisker-evoked Ca^{2+} and Na^+ action potentials in distal and terminal dendrites of rat neocortical pyramidal neurons *in vivo*. *J. Neurosci.* **22,** 6991–7005 (2002).
38. Connors, B. W. & Prince, D. A. Effects of local anesthetic QX-314 on the membrane properties of hippocampal pyramidal neurons. *J. Pharmacol. Exp. Ther.* **220,** 476–481 (1982).
39. Mooney, R. Synaptic basis for developmental plasticity in a birdsong nucleus. *J. Neurosci.* **12,** 2464–2477 (1992).
40. Dave, A. S. & Margoliash, D. Song replay during sleep and computational rules for sensorimotor vocal learning. *Science* **290,** 812–816 (2000).
41. Hahnloser, R. H., Kozhevnikov, A. A. & Fee, M. S. Sleep-related neural activity in a premotor and a basal-ganglia pathway of the songbird. *J. Neurophysiol.* **96,** 794–812 (2006).
42. Wei, D. S. *et al.* Compartmentalized and binary behavior of terminal dendrites in hippocampal pyramidal neurons. *Science* **293,** 2272–2275 (2001).
43. Mooney, R. & Prather, J. F. The HVC microcircuit: the synaptic basis for interactions between song motor and vocal plasticity pathways. *J. Neurosci.* **25,** 1952–1964 (2005).
44. Jun, J. K. & Jin, D. Z. Development of neural circuitry for precise temporal sequences through spontaneous activity, axon remodeling, and synaptic plasticity. *PLoS ONE* **2,** e723 (2007).
45. Fiete, I. R., Senn, W., Wang, C. Z. & Hahnloser, R. H. Spike-time-dependent plasticity and heterosynaptic competition organize networks to produce long scale-free sequences of neural activity. *Neuron* **65,** 563–576 (2010).
46. Margrie, T. W., Brecht, M. & Sakmann, B. *In vivo*, low-resistance, whole-cell recordings from neurons in the anaesthetized and awake mammalian brain. *Pflugers Arch.* **444,** 491–498 (2002).